Estimation of the size of informal economy in Russian Federation based on household budget survey data

Yaroslav Murashov

▶ **To cite this version:**

Yaroslav Murashov. Estimation of the size of informal economy in Russian Federation based on household budget survey data. Economics and Finance. 2014. dumas-01095923

HAL Id: dumas-01095923

https://dumas.ccsd.cnrs.fr/dumas-01095923

Submitted on 16 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Paris1 – UFR 02 Sciences Economiques

Master 2 Recherche Economie Théorique et Empirique

"Estimation of the size of informal economy in Russian Federation based on household budget survey data"

Nom du directeur de la soutenance: Francois Gardes

Présenté et soutenu par : Murashov Yaroslav

Paris, 2014

L'université de Paris 1 Panthéon Sorbonne n'entend donner aucune approbation, ni désapprobation aux opinions émises dans ce mémoire ; elles doivent être considérées comme propre à leur auteur

CONTENTS

CONTENTS.....3
THE DEFINTION AND METHODS TO IDENTIFY THE INFORMAL ECONOMY.....4
RLMS DATABASE DESCIPTION.....9
INCOME DECOMPOSITION.....14
EXPENDITURE DECOMPOSITION.....21
ECONOMETRIC MODEL ESTIMATION.....29
APPLICATION TABLES.....43
REFERENCES.....47

The definition and methods to identify the informal economy.

Black economy is by definition the activity concealed from law. So it will not be indicated in the official revenue documents. Although, if we are looking at the revenue surveys, they may include some information about the black economy activities, but to some extent, because people still have some incentives to conceal black economy activities. Therefore, researches may try to estimate the part of the economic activity which is treated as the black economy. One of the ways to estimate the black economy is microeconomic approach and is based on the data obtained from households budget surveys. The key idea is that the households indicate their expenditures more accurately than they indicate their income.

Actually, the main idea of the authors Pissarides and Weber, 1989 [3] is that 1) the reporting of expenditure on *some items* by *all groups* of population is accurate; 2) the reporting of income by *some groups* of the population is accurate;. Actually, the income reporting is accurate not by some groups, but by some types of occupation of the population. The authors believe that the expenditure item which is recorded correctly is the expenditures on food (the less likely to conceal). The underreport of income comes from the people who are self employed. Employees report their income correctly. Although this may look like a strong assumption that only the self - employed are those who under - report their income, because for most of the people to indicate the wage expenditures is not always reasonable. They may obtain other wages which are not indicated in the responses to the questionnaire. That is why it may be important to use the variable of total income earned by the household, which is not divided into subgroups and may be used as the base for calculation of self - employment income.

Therefore, the authors propose straightforward method to estimate the size of black economy. First, an expenditure function is needed to be calculated. Then the expenditure function is inverted and the income is forecasted from the reported expenditure.

First of all it is important to build a theoretical model, describing household consumption patterns, to account for the underreport of income on the one hand, and the connection between income and consumption, on the other hand.

The variables reported by the households: 1) consumption of individual items C_{ij} (i is the household index, j is the index of the index of the consumption item); 2) after - tax income $Y_i^{'}$; 3) vector of household characteristics Z_i. According to our assumptions, there is underreport of the level of income for households in self - employment. Denote by k_i the coefficient by which a household i is underreporting their income. Then the connection between the factual and reported income is written in the following form: $Y_i = k_i * Y_i^{'}$. So $k_i = 1$ for employees and $k_i \geq 1$ for the self - employed. The expenditure function of item j is written in the following form: $lnC_{ij} = Z_i\alpha_j + \beta_j lnY_i^p + \varepsilon_{ij}$. It is needed to note the definition of Y_i^p. We denote by this the measurement of the income influencing consumption decisions. The relation between the permanent and actual income is measured by $Y_i = p_i * Y_i^p$. This parameter accounts for the variation of income due to the unforeseen circumstances. The mean of p_i does not depend on the type of the household and is the same for the employees and the self - employed. On the other hand, the variation of the parameter is different depending on the type of a household: $Var(p_i|employee) < Var(p_i|self - employed)$. If the household is self - employed, then the variation of this parameter is higher, reflecting the higher variation of the self - employed income. That is why current consumption is a function of not a current income, but the permanent income.

Therefore, permanent income, which is directly related to the consumption function, can be decomposed into current income and the parameters of the model through the following way:

$$lnY_i^p = lnY_i^{'} - lnp_i + lnk_i$$

This implies existence of the two additional random regressors in the model if we put into the model observed income instead of the permanent income. To verify statistical hypothesis the authors must make assumptions concerning the statistical distribution of the coefficients responsible for the income mismatch. The coefficients have log-normal distribution. So now write them in the form of deviations from their means. $lnp_i = \mu_p + u_i$; $lnk_i = \mu_k + v_i$. Now the trick performing the connection between the mean of p_i and the mean of its log: $ln\bar{p} = \mu_p + \frac{1}{2}{\sigma_u}^2$. As far as the mean of p_i does not depend on the type of the household, one may compare the mean of the log for the employees and the self - employed: $\mu_{pse} - \mu_{pee} = \frac{1}{2}({\sigma_{ee}}^2 - {\sigma_{se}}^2) \leq 0$. How one can use this fact? Let us put the error terms decomposition into the expenditure function. Then we obtain the following decomposition of permanent income: $lnY_i^p = lnY_i^{'} - (\mu_p + u_i) + (\mu_k + v_i)$. This decomposition can be used for the expenditure function. Then:

$$lnC_{ij} = Z_i\alpha_j + \beta_j lnY_i^{'} - \beta_j(\mu_p - \mu_k) - \beta_j(u_i - v_i) + \varepsilon_{ij}$$

The dependent variable is the household expenditures on some particular item. In authors model food expenditures will be considered. The idea of this quite simple model is to obtain income differences for the self - employed. If this equation is estimated separately for employees and self - employed then the intercepts differ as far as $(\mu_p - \mu_k)$ is not the same at each group. This constant term may give information on the size of black economy.

To estimate this relationship the following type of regression model was used:

$$lnC_{ij} = Z_i\alpha_j + \beta_j lnY_i^{'} + \gamma_j SE_i + \mu_i$$

Where SE_i is a dummy variable taking value "1" if the individual is self - employed, and zero if he is employee. This model can be straightforward estimated by the method of least squares, accounting for the heteroscedasticity of error term μ_i. But how can we exactly compute the level of income under - reporting with the help of this model?

According with the theoretical income decomposition for the expenditure function the coefficient γ_j must be equal to the term $\{-\beta_j(\mu_p - \mu_k)|self - employed\} - \{-\beta_j(\mu_p - \mu_k)|employee\}$. We can see that our hypothesis for the equal coefficient β_j for both self - employed and employees holds. This enables us to perform the model. From our previous discussion $\mu_{pse} - \mu_{pee} = -\frac{1}{2}({\sigma_{use}}^2 - {\sigma_{uee}}^2)$; whereas $\mu_{kse} - \mu_{kee} = \mu_{kse} = \mu_k$. Therefore $(\mu_{pse} - \mu_{pee}) - (\mu_{kse} - \mu_{kee}) = -\frac{1}{2}({\sigma_{use}}^2 - {\sigma_{uee}}^2) - \mu_k$.

The estimated coefficient equals: $\gamma_j = (-\beta_j) * (-\frac{1}{2}({\sigma_{use}}^2 - {\sigma_{uee}}^2) - \mu_k) = \beta_j[\mu_k + \frac{1}{2}({\sigma_{use}}^2 - {\sigma_{uee}}^2)]$.

The income decomposition enables us to perform the reduced - form regressions for income. Therefore the observable income can be decomposed in the following form:

$$lnY_i^{'} = Z_i\delta_1 + X_i\delta_2 + \eta_i$$

Due to the income decomposition, we obtain that $Var(\eta_{se}) - Var(\eta_{ee}) = Var(u - v)_{se} - Var(u)_{ee}$; or equivalently: ${\sigma_{Yse}}^2 - {\sigma_{Yee}}^2 = {\sigma_{use}}^2 - 2cov(uv)_{se} + {\sigma_{vse}}^2 - {\sigma_{uee}}^2$.

Important implication of this decomposition is as follows: for given value of ${\sigma_{uee}}^2$, ${\sigma_{vse}}^2$ and ${\sigma_{use}}^2$ are negatively related.

From the estimated coefficient γ_j we have that $\mu_k = \frac{\gamma_j}{\beta_j} - \frac{1}{2}({\sigma_{use}}^2 - {\sigma_{uee}}^2)$. Then we have $\mu_k + \frac{1}{2}{\sigma_{vse}}^2 = \frac{\gamma_j}{\beta_j} + \frac{1}{2}({\sigma_{vse}}^2 - {\sigma_{use}}^2 + {\sigma_{uee}}^2)$.

We get the lower bound for mean under - reporting if ${\sigma_{vse}}^2 = 0$; the upper bound of mean under - reporting is obtained when ${\sigma_{use}}^2$ is minimized. Having the income variance decomposition and assuming that the error terms u and v are not correlated we can determine the bounds for the level of income under - reporting. They are as follows:

$$lower\ -\ bound: \overline{lnk} = \mu_k + \frac{1}{2}{\sigma_{vse}}^2 = \frac{\gamma_j}{\beta_j} - \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$$

$$upper\ -\ bound: \overline{lnk} = \mu_k + \frac{1}{2}{\sigma_{vse}}^2 = \frac{\gamma_j}{\beta_j} + \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$$

With the help of this decomposition the interval for the level of income under - reporting can be obtained, since we have estimated the model of expenditures and computed the income variations.

The estimating procedure for this is quite straightforward:

1) we pick some expenditure item and monitor the households expenditure on it (the expenditures on the item should be reported correctly);

2) denote the type of the household (either a self - employed or the employee), according with the source of the main income (classified as self - employed if the income from self - employment is not less than 25% of total income);

3) estimate the model in the following form: $lnC_{ij} = Z_i\alpha_j + \beta_j lnY_i^{'} + \gamma_j SE_i + \mu_i$, the aim is to estimate the marginal propensity to consume (β_j) and the parameter (γ_j);

4) estimate the income variance depending on the type of the household;

5) compute the lower and upper bound of the level of income under - reporting, assuming zero covariance between the errors in the model.

In the article written by Lyssiotou and Pashardes [2] the authors actually use the method of Pissarides and Webber [3] proposed in 1989 to estimate single food expenditure equation. But the authors go further n the estimation proposing the two other methods of black economy estimation.

The basic way of developing the model of Pissarides and Webber used by the authors is to include in the analyses the demand for durable goods. By dividing the demand into the demand for durable and the demand for non-durable goods one may notice that the share of income a household spends on durable goods depends not only on the level of income, but on the income source. Therefore, self - employment income may indicate not only income under-report, but the preference heterogeneity: self - employed people may tend to spend more on expensive goods (consumption of durable goods), therefore they are spending too little on food and other non-durable goods.

Another reason why the self - employment income is not a good proxy for the utility is that it tends to be more volatile, therefore, it influences savings. Households spend less and earn more than employees to meet the precautionary savings.

The authors separate the preference structure so that the consumption is divided between durable and nondurable goods. The cost functions of consumption are defined: $C(p; U) =$

$F[c(p;U);d(r;U);U]$, where $c(.),d(.)$ denote the sub-cost functions. Using this function one can depict household expenditure on the $i-th$ good:

$$p_i q_i = p_i \frac{\partial F[.]}{\partial c(.)} \frac{\partial c(.)}{\partial p_i} = p_i y \frac{\partial c(.)}{c(.)\partial p_i}$$

The budget share of good i in the expenditures on nondurable goods:

$$w_i = \frac{\partial lnc(.)}{\partial lnp_i}$$

The authors impose the unit cost of nondurables $c(.)$. It has a quadratic logarithmic form:

$$lnc(p;U) = a(p) + b(p)[\frac{U}{1-g(p)U}]$$

Using this function by differentiating one can obtain Hiksian shares of the demand (w_i) denoted above. This can be parameterized with the following function:

$$w_i = a_i(p) + b_i(p)[\frac{U}{1-g(p)U}] + \lambda_i(p)[\frac{U}{1-g(p)U}]^2$$

Given the cost function we can see that the expression $\frac{U}{1-g_o U}$ is directly linked to the level of income of a household through the indirect utility function, so authors make linear decomposition of this expression as a function of lnY^*, where Y^* is the true household income. As the result Hiksian demand shares can be written as:

$$w_{ih} = a_i + \beta_i(lnY_h^*) + \lambda_i(lnY_h^*)^2$$

where the parameters are a function of Hiksian demand shares parameters, the price index is dropped, since the prices are fixed at the level $p_i = 1, r_i = 1;\ all\ i$. The label h denotes a household, the label i denotes a good in the consumer bundle. This is the theoretical foundation for the Engel curve, since the consumption share depends on income as a quadratic function. The quadratic form of the Engel curve means that since the income of a household increases it tends to spend more expenditure share on luxury goods and less expenditure share on necessities.

The advantage of this model is that there is no need to arbitrary impose the type of a household. This can be seen through the following income decomposition:

$$Y_{kh}^* = \theta_k Y_{kh}, where\ \theta_k\ -\ kth\ component\ of\ income$$

The share of income earned by a household from a certain income source is computed, this is denoted by $y_{kh} = \frac{Y_{kh}}{Y_h}$. The total real income of a household equals $Y_h^* = \sum_k \theta_k Y_{kh}$. Dividing both part of the equation by the observable income: $\frac{Y_h^*}{Y_h} = \frac{\sum_k \theta_k Y_{kh}}{Y_h}$; therefore $\frac{Y_h^*}{Y_h} = \sum_k \theta_k y_{kh} \Longrightarrow Y_h^* = (Y_h) * (\sum_k \theta_k y_{kh})$. This helps to rewrite the Engel curve equation:

$w_{ih} = \alpha_i + \beta_i[lnY_h + ln(\sum_k \theta_k y_{kh})] + \lambda_i[lnY_h + ln(\sum_k \theta_k y_{kh})]^2$i

1) The model is more complex than one considered by Pissarides and Weber, 1989. But it can be linked to the model of Weber if one can impose a certain threshold in the share of household's self - employment income:

y_{kh} is replaced by $\begin{cases} D_{kh} = 1\ if\ y_{kh} > \widetilde{y_{kh}} \\ D_{kh} = 0\ othewise \end{cases}$

λ_i equals zero for all i

The equation for the commodity i according to Pissarides, Weber looks like:

$$w_{ih} = \alpha_i + \beta_i lnY_h + \sum_k \beta_i\, ln\theta_k D_{kh}$$

This equation is estimated for the category of goods for which the income is reported the most correctly. This category is food expenditures.

Actually in the article the following model is estimated, almost similar to Pissarides, Weber:

$$lnY_{Fh} = a_F + \delta_F S_h + \beta_F lnY_h + \lambda_F (lnY_h)^2 + \sum_j \alpha_{Fj} z_{jh} + u_{Fh}$$

The problem is that the expenditure on the left - hand side is actually varies along with not only the total income, but with the type of the household. Thus reflecting the preference heterogeneity:

$$\frac{\Delta w_{ih}/\Delta D_{kh}}{\Delta w_{jh}/\Delta D_{kh}} = \frac{\Delta w_{ih}/\Delta lnY_h}{\Delta w_{jh}/\Delta lnY_h}$$

The single equation approach does not distinguish these effects, that is why it is limited. The system approach can cope with this difficulty and is performed by the authors.

2) Actually the first way to estimate the black economy complementary to the Pissarides, Webber approach is to use the nonparametric approach. There exists a nonlinear function which determines the Engel curve equation for the household of each type. The idea is to estimate the nonparametric regression for each type of the household (employee, self-employed) and to measure the distance between the expenditures through the nonlinear part of the model:

$$Y_{Fh}^l = c_i^l z_h + m(Y_h^l) + u_h^l, where\ l\ is\ the\ type\ of\ a\ household$$

3) The third method used by the authors is the demand system approach. As have been mentioned above, this method enables to account for both heterogeneity of preferences and income under - report. As in Pissarides and Weber, households are assumed to have two sources of income: income from wage and income from self - employment. But there is no a certain discrete type of a household: in the model authors use budget shares of the household income. The income under-report parameter does not include preference heterogeneity, since the heterogeneity is represented by a term shifting the Engel curve. The demand system is based on the household expenditures shares for non - durable goods. The following good categories are considered: food, alcohol, clothing, personal goods, leisure goods, fuel. The equation to be estimated is written in the following form:

$$w_{ih} = \alpha_i + \sum_j \alpha_{ij} z_{hj} + \delta_i y_h^s + \beta_i [lnY_h + ln(\theta_0 y_h^w + \theta_1 y_h^s)] + \lambda_i [lnY_h + ln(\theta_0 y_h^w + \theta_1 y_h^s)]^2 + v_{ih}$$

There are three major differences of this model in comparison with Pissarides and Webber, 1989:

1) The authors estimate a system of budget share equations (the good is denoted by i);

2) They use not the dummy variables indicating the role of household occupation, but the share of income obtained from a certain activity;

3) The parameter δ_i is no longer responsible for the level of income under-report, but stands for the preference heterogeneity.

Estimation of the size of informal economy in Turkey estimated by Aktuna Gunes A., Starzec C., Gardes F. [1] is based on the model proposed by Lyssiotou and Pashardes, 2004. But now the authors assume that a household has three forms of income: wage income, self - employment and other income. The other income seems to be reported correctly, while the other income components may be hidden. Due to the fact $\theta_k \geq 1$, for $k - wage\ income, self -$

employment income. Another enlargement is that the preference heterogeneity term depends on the share of each part of the income (the heterogeneity is viewed in the form of difference in savings between different occupation groups and consumption patterns).

The estimated system of equations is actually the same. The difference is that the system is estimated by generalized method of moments and the household production activities are included in the system in the form of household consumption and household production. So the initial model is enlarged and includes the consumption of household goods and the income from their production evaluated by the market price.

However, estimating the size of informal economy in Russian Federation we are not able to compute the value of household production since we do not know the amount of time used on a certain activity (there is no such a question in the interview neither on the household nor on the individual level). But there is the information on the household production of the agricultural goods. The money income from the goods sold is actually included in the household total income. The information of goods consumed is added to the goods consumption data to form full expenditures. The value of the household goods produced is computed on the basis of the purchase prices indicated in the interview paid for a certain type of commodity. Therefore, it represents the market valuation of household production.

On the first stage it is possible to estimate a single food expenditure equation and to see whether an income decomposition and the black economy coefficient can be obtained. To say in advance, there are some problems concerning both the income decomposition and the estimation of this coefficient. Therefore, the more complicated econometrician estimation methods should be used.

RLMS database description

Russian Longitudinal Monitoring Survey is the only annual nongovernment monitoring of social - economic characteristics of the population of Russian Federation and the health conditions [4].

Monitoring represents a series of representative questions based on the multi - level probability multi-step territorial sample.

The key peculiarity of RLMS is the wide - spread base of socio - economic variables. The variables include the income and expenditure structure, material welfare, investment, occupation, migration, health conditions, structure of food consumption, education. RLMS is denoted by panel data, modern methodology, data comparability.

What is special about the survey?

- The methodology of the questionnaire enables world - wide comparisons;

- Thanks to RLMS for the first time the researches got access to the alternative to government statistics microdata on the base of national sample;

- The monitoring program contains some important characteristics which are not available in the government statistics. It is the only survey containing the initial information on *income of certain household members and the households as a whole*

- RLMS is practically the only representative microeconomic survey in Russian Federation having a large panel component: the same households are interviewed during a long period of time; this increases the quality of forecast based on RLMS;

- RLMS contains a large block of valuating questions which add to the information about the change in Russian households life activity by the subjective assessment of changes in the country.

RLMS has been conducted during the two periods: the first period survey has been made from 1992 to 1993; the second - since 1994 till now. All the work for the monitoring including the sample design, has been made by the Russian research team supervised by Kozireva P. M. and Kosolapov M. S.

Starting from 2006 National Research University - Higher School of Economics is playing a key role in the project.

The model of sample The dynamic aspects of the analysis of changes in the households claims "panel" sample model, where in each wave the same household members or individuals are questioned. But with the time this panel sample is growing older, certain elements disappear and systematic bias arises. Therefore the panel model does not allow for a clear situation at the moment of the new wave. This task is accomplished by cross - sectional model of sample.

The cross-sectional data is combined with the panel. The model was chosen such that along with cross-sectional sample enables to provide panel analysis. This solution is named repeated sample with split - panel. The advantage of the repeated sample is that it provides data analyses on both the households and the household members.

The data collection periods of RLMS include 21 waves. The last one has been conducted from October 2012 to December 2012.

Our attention is devoted to the study of households of cross-sectional analyses - the households which represent the population of Russian Federation and actually live at the adresses included in the survey.

The identification variables are used to identify households and individuals in different waves. In each wave besides the identification variables of the same wave there are also identification variables of the preceding waves.

The first part of the research will be devoted to the analyses of representative sample for the household in the last wave of the research. The number of this wave is 21. The RLMS household budget survey has been conducted in this wave from October 2012 to December 2012. We have the representative sample which represents the population of Russian Federation at the moment the survey has been conducted. It means that only the individuals which are available at the place of their living are interviewed. That is why the sample represents the current situation in the household, and, as far as the sample is representative, the current situation in Russian Federation.

The wave contains the file of individual data and the file of household data. The file of individuals contains unique individual identification number for each respondent. Also it contains the number of household the member of which the individual is during the wave. This linked to the household number in the file of the household. As far as the unit of observation in our analyses is a household, this link from individual to household is used to merge the data from individual database with the data from the household database.

At the first stage, this database merging must be made. It is performed by sorting individuals by family number and by the variable of interest. Then the needed variable is kept and added to the household database. The following individual variables are used:

1. Family status:

D_couple=1: an individual is in a registered marriage (we have a family) (_marst==2);

D_widow=1: an individual is a widow (_marst==5);

D_divorced=1: an individual is divorced (_marst==4);

The basic group of a family status is lonely or not registered relationships.

2. Occupational status (if any member of a household has a given occupation):

D_white_collar=1: occupational status of an individual is one of the following - 1) law-makers or government workers (_occup==1), 2) specialists of higher qualification level (occup==2), 3) qualified agricultural workers (_occup==6), 4) qualified workers of hand - work (_occup==7), 5) other qualified workers (_occup==8). The basic group is blue collar workers, therefore blue collars are occupied in one of the following spheres: specialists of average qualification level, clerks, workers in the sphere of trade and service;

D_white_collar_male=1: occupational status of the head of the family is white collar (D_white_collar==1) & (D_male==1~h5==1);

3. Level of education (if any member of a household has a certain education degree):

D_educ_sec_special=1 finished special education (_diplom==5);

D_educ_higher=1 finished higher education level (or higher) (_diplom==6);

The basic educational group is the group of either full or not full secondary education.

4. The level of life satisfaction (if any member of a household reports that he is fully satisfied with life)

D_life_satisfaction=1 (j65==1);

5. The presence of some stomach deseases: D_stomach_desease=1 (m20_65==1). This variable may be useful when estimated household demand for food;

6. The desire to find another job: D_wish_other_job=1 (j81==1). This variable indicates whether the household may be willing to participate in informal economy activities;

7. The children in the household: D_children=1 (j72_171==1);

8. The number of children in the household: D_number_of_children (j72_172);

9. The ability of a household to improve the living conditions: D_living_improvement (j721631==1);

10. The ability of a household to have a vacation with all members of the family: D_vacation_possible (j721634==1);

11. The ability of a household to pay for a child study in the University: D_child_study_pay (j721635==1);

12. The presence of some other job at one member of the household: D_have_other_job (j32==1).

So what is the major difference between the characteristics of individuals in the households with no income from self - employment and those with high level of self - employment income? The comparison of individual characteristics of the households can be seen in the Table 1.

Individual characteristics descriptive statistics				
Variable	% total	% non_se	% 0-20_se	% 20-100_se
D_couple	52.80%	54.09%	52.54%	48.28%
D_white_collar	45.80%	49.79%	42.12%	38.17%
D_white_collar_male	33.32%	35.52%	32.00%	27.53%
D_educ_sec_special	41.76%	42.49%	42.32%	37.74%
D_educ_higher	36.69%	39.05%	33.80%	33.55%
D_children	89.87%	90.05%	90.70%	87.42%
D_life_satisfaction	15.51%	15.29%	15.15%	17.10%
D_wish_other_job	31.19%	29.71%	29.94%	39.57%
D_have_other_job	4.24%	4.60%	3.54%	4.30%
D_living_improvement	11.05%	11.63%	9.09%	12.90%
D_vacation_possible	20.57%	22.68%	17.72%	18.28%
D_child_study_pay	19.27%	20.53%	16.74%	19.68%

Table 1. Individual characteristics descriptive statistics

Considering the family status self-employed are usually single than married. They are more involved in the non - qualified labor, both for the head and other members. They have lower level of education: both higher education (33.55% vs 39.05%) and secondary special education (37.74% vs 42.49%). They report a bit more higher level of life satisfaction (17.10% vs 15.29%). The greater part of them does wish other job (39.57% vs 29.71%). But they do not in general indicate that they have other job (maybe they conceal their shadow part of income, or simply do not have other job for which to be paid). To speak of financial possibilities, their possibilities to improve living conditions, have vacation, pay for child study do not differ from the average sample level and the level of non self-employed but exceed the possibilities of those who have minor share of self employment income. Therefore some of these characteristics may be very useful to account for the structural differences between different household types and used in regression analyses.

All these variables represent individual characteristics but are attributed to the households. They are used in the regression model to reflect the differences in preferences between the individuals. The choice of variables comes mainly from the database restrictions and the representation of key socio - economic characteristics. Also the variables include the questions giving information about the possibility of households to spend money on various activities.

The part of the survey devoted to the households also has the variables characterizing households but not the income or expenditure components.

These are the questions concerning:

1. The place of living: type of town, population.

2. Living conditions: type of house, price, home owner, total area, living area, number of rooms, central heating, central water, hot water, canalisation, telephone, magistral gas, electro owen, refrigerator, frost, auto washing machine, micro wave owen, dish washing machine, colour TV, TV plazma, videopleer, DVD_pleer, computer, notebook, low speed internet connection, digital camera, video camera, MP3 pleer, GPRS navigator, home auto, import auto, lorry, motor cycle, bicycle, tractor, lawn mower, air conditioner, sputnic antenna, cabel TV, garden house, other appartment.

3. Some variables concerning land use: have you got any land at use; the land area; has you paid for land; have you grown something, have you sold something.

4. The educational questions: do children go to school; does a household member visit college.

Here you can see (Table 2) the descriptive statistics of these variables with respect to income groups subsamples (the decomposition of income see at the income decomposition description section):

Condition	Variable	% total	% non_se	% 0-20_se	% 20-100_se
status==4	Village	24.0%	21.2%	24.8%	33.2%
status==3	small town	5.6%	5.9%	5.2%	5.3%
status==2	Town	26.3%	27.4%	27.7%	18.7%
status==1	City	44.1%	45.4%	42.3%	42.8%
Popul (mean)	population	1541327	1590014	1468026	1504747
c1==1	your_house	88.9%	89.6%	91.3%	81.4%
c1==2	D_rent_house	8.8%	8.6%	7.0%	13.3%
c1==3	D_common_house	2.2%	1.7%	1.7%	5.3%
c1_1 (mean)	dwelling price	1923205	1973695	1856279	1866193
c4_0==1	self_owned	90.1%	91.0%	89.0%	89.0%
c4_0==2	D_home_relative_owned	3.9%	3.6%	3.7%	5.3%
c4_0==3	D_home_not_privatised	6.0%	5.4%	7.3%	5.7%
c6	total_square	54.9	54.9	54	57.1
c5	living_square	36.7	36.4	35.8	39.5
c5_1	number of rooms	2.3	2.3	2.3	2.4
c7_1==1	D_cental_heating	71.7%	74.6%	70.4%	62.9%
c7_2==1	D_central_water	87.8%	88.7%	87.0%	85.6%
c7_3==1	D_hot_water	66.2%	68.5%	65.4%	58.7%
c7_5==1	D_canalisation	73.7%	76.8%	72.3%	64.3%
c7_6==1	D_phone	59.1%	59.3%	63.4%	49.2%
c7_7==1	D_magistral_gas	66.0%	67.9%	65.5%	60.0%
c7_8==1	D_owen	23.4%	24.2%	21.5%	24.3%
c9_1_1a==1	D_refrigerator	49.5%	51.3%	46.8%	48.0%
c9_2a==1	D_frost	11.9%	12.4%	11.1%	11.5%
c9_3_2a==1	D_washing_machine	72.3%	74.9%	68.6%	69.9%
c9_3_1a==1	D_micro_wave	61.2%	63.0%	58.9%	58.5%
c9_3_3a==1	D_dish_washing	2.7%	3.1%	1.9%	3.0%
c9_5_1a==1	D_colour_TV	78.5%	77.9%	81.2%	75.7%
c9_5_2a==1	D_plazma	42.3%	44.7%	40.2%	37.3%
c9_6a==1	D_pleer	24.7%	26.3%	23.2%	21.9%
c9_6_0a==1	D_DVD	45.5%	45.1%	44.3%	49.0%
c9_621a==1	D_computer	42.1%	44.6%	39.2%	38.5%
c9_622a==1	D_notebook	32.3%	33.9%	28.1%	34.7%
c9_623a==1	D_low_speed_int	18.1%	18.8%	16.3%	19.5%
c9_624a==1	D_high_speed_int	37.0%	38.8%	34.2%	36.0%
c9_6_3a==1	D_digital_camera	36.9%	38.8%	33.6%	36.5%
c9_631a==1	D_video_camera	7.7%	7.8%	7.6%	7.4%
c9_6_4a==1	D_MP3	10.4%	11.1%	8.8%	11.0%
c9_6_5a==1	D_GPRS	7.2%	8.1%	6.1%	6.5%
c9_7_2a==1	D_home_auto	22.0%	21.4%	23.2%	21.6%
c9_7_3a==1	D_import_auto	20.7%	23.3%	17.3%	17.6%

c9_7_1a==1	D_lorry	2.3%	2.3%	2.0%	2.9%
c9_8a==1	D_motor_cycle	3.1%	2.9%	3.2%	3.7%
c9_8_1a==1	D_bicycle	21.2%	20.1%	23.2%	20.9%
c9_9a==1	D_tractor	2.4%	2.1%	2.3%	3.7%
c9_9_1a==1	D_lawn_mover	7.5%	8.0%	7.5%	5.2%
c9_13a==1	D_air_conditioner	8.2%	8.1%	8.2%	8.3%
c9_14a==1	D_sputnik_antenna	17.0%	15.5%	17.4%	21.9%
c9_15a==1	D_cabel_TV	30.5%	32.4%	29.5%	25.1%
c9_101a==1	D_garden_house	21.3%	22.6%	21.6%	15.3%
c9_12a==1	D_other_appartment	8.5%	8.6%	7.9%	9.7%
d1==1	D_land_use	50.1%	49.1%	52.2%	49.4%
d5==1	D_paid_land_use	34.2%	34.0%	34.6%	34.2%
d7==1	D_grown_smth	44.8%	43.4%	47.7%	44.2%
d9==1	D_products_sold	2.1%	1.1%	2.2%	5.7%
h21==1	D_secondary_education	30.4%	28.5%	30.6%	37.6%
h45==1	D_college	16.8%	17.1%	14.7%	19.9%

Table 2. Descriptive statistics of household characteristics

Actually, there is quite little information to take from this descriptive statistics table. The percent values define the share of people in the sample which own a certain durable good or have a certain occupation. Most of the mean values of the variables do not differ from one occupational group to another. There can be made no conclusion that self - employed have a greater amount of durable goods and therefore that their wealth is higher. Maybe that is due to the fact that the ownership of these goods are not reported in the interview.

Despite this, some conclusions can be made from this table. The self - employed tend to be more settle in the village (33.2% vs 24.8% and 21.2% non_se), and also have more share of people which sell their products grown (5.7% vs 2.2% and 1.1% non_se). Less part of self - employed have phone (49.2% vs 63.4% and 59.3% non_se) Therefore, maybe some part of their income may be the agricultural income from the selling of products. Also as we have mentioned in the literature review the agricultural goods consumption shall be included is household foods product consumption to provide unbiased value of food consumption. This is the point to correct our model.

Income decomposition

An important part of our study is to see how income is decomposed between the different sources among the households. First of all, the total income of the household consists of three parts: 1) wage income free of taxes; 2) the part such as other income usually fixed income; 3) income from self-employment. The last one is actually computed as a difference between total income of a household, the wage income and the income from all other sources (which form non_wage income in our model, more precisely other income). So the formula by which the income is built is looking as follows:

$$self\ employment\ income = total\ income - wage\ income - other\ income$$

$$non\ wage\ income = total\ income - wage\ income$$

The non_wage income of a household includes: pension; scholarship; unemployment benefit; income from equity sold; income from rent of equity; capital income in the form of the interest; capital income in the form of the dividend; insurance premium; aliments; money from the debt reimbursement; subsidies from the appartment payment. Actually the inclusion in this form the income from equity sold does not prove to be reasonable, because it is not a permanent source of income and therefore may be misleading. For the first time let us distinguish between non_wage income and other income. First is simply the difference between total income and wage income. While "other income" is free of self - employment income.

Now let us introduce some descriptive statistics to characterize the variables of interest. Here are the descriptive statistics of non_wage part of the income (Table 3). As you can see in the descriptive statistics, the most popular item indicated by the households is pension, indicated by 3910 households out of 6517 in the total sample. As an outlier stands the income from equity sold. There are quite few observations, but they account for the maximum value of one million and seven hundred thousand rubles. Therefore we do not include this form of income in the household sources of income.

Speaking about the income from capital (to the capital income we must denote the income from rent of equity, capital income in the form of interest and capital income in the form of dividend) we notice that there is very low rate of response (only 150 people have indicated that they have got income in some of the forms). This can be explain twofold: 1) most of respondents conceal the real amount of capital income because of the fact that they do not want to give any information about the total amount of capital owned; 2) the capital forms of income are not so popular in our country with comparison to the households in the western countries. Anyway, there is a reason not to believe these figures and consider that there is something behind them. We must say that non_wage income computed using the information on capital income is under - reported. Therefore, the self - employment income as it is defined must include some forms of capital income. So both parts: capital income which is fixed and less variable then self - employment income are under - reported and must be estimated. Although while the former is concealed, it is not mandatory that it is concealed for the reason of tax evasion and therefore should be viewed as a part of black economy.

	Variable names				
stats	Pension	scholarship	unemployment benefit	income from equity sold	income from rent of equity
N	3910	287	131	77	105
Mean	13786.81	1815.129	2996.412	139027.3	11952.38
Min	1030	300	400	500	2000
Max	64658	13000	80000	1700000	50000
p25	8500	600	1000	7000	4500
p50	11600	1200	2000	18000	9000
p75	17775	2000	3000	48000	15000

	Variable names					
stats	capital income in the form of interest	capital income in the form of dividend	insurance premium	aliments	money from the debt reimbursement	subsidies from the apartment payment
N	43	5	9	184	164	1592
Mean	6717.86	11320	23874.67	5820.625	6429.055	975.7469
Min	20	400	1800	300	50	50
Max	50000	40000	120000	35000	150000	6800
p25	1000	1200	7000	2500	1000	500.5
p50	2000	5000	10000	4000	2000	800
p75	7500	10000	20000	6600	5000	1200

Table 3. Indicated parts of non wage income

The sources of income which are combined under the name "other income" represent the following items: pension; scholarship; unemployment benefit; income from rent of equity; capital income in the form of the interest; capital income in the form of the dividend; insurance premium; aliments; money from the debt reimbursement; subsidies from the house payment. We exclude income from equity sold from the items including in "other income". The wage income, other income and total income provide the basis for the computing of self - employment income. "non_wage income" represents the difference between total income and wage income. The description of household income decomposition can be seen in Table 4.

	Variable names				
Stats	wage_inc	total_inc	other_inc	non_wage_inc	self_employment_inc
N	4412	6215	4475	4844	3786
"zero" value	2105	302	2042	1673	2731
Mean	33699.99	36547.41	13477.98	16196.9	4792.341
Min	990	400	85	-375400	-381000
Max	420000	425300	175000	332300	320300
p25	15000	16000	7800	7614	100
p50	26000	27600	11200	12359	1448.5
p75	44000	45430	17800	20205	5917

Table 4. Descriptive statistics of income decomposition

As one can see from this table, positive total income is indicated by the most of households, only 302 do not indicate their income. For the analyses we do not distinguish between "zero" value of the variable and the "missing" value, since for the wage income those who report "zero wage" are added to those who report "I do not get any wage income". So a considerable part of the sample (2105) say they do not get any wage. These may be pensioners for whom pension is the only source of income.

Now we turn to the description of non_wage income: as we can see "zero" value of non_wage income is attributed to 1673 households. As far as this is simply the difference between total income and wage income, one may suppose that this is simply the amount of the households for which all the income is formed by wage. Another interesting characteristics of non wage income is that some have negative non_wage income, as well as self employment income. It means that for the households the total income is indicated incorrectly and it must be replaced by the sum of wage income and other income . We make the replacement for those households for which the value of non_wage income and income from self_employment is negative. It means automatically that self_employment income for those households becomes equal zero. But for a moment we will denote the households with the "wrong" self - employment income as a special part of a sample.

The corrected descriptive statistics table (Table 5) for the income decomposition looks as follows:

	Variable names				
Stats	wage_inc	total_inc	other_inc	non_wage_inc	self_employment_inc
N	4412	6359	4475	5241	2880
"zero" value	2105	158	2042	1276	3637
Mean	33700	37126	13478	16676	9405
Sd	29838	33929	9532	18867	21905
sd/mean	0.89	0.91	0.71	1.13	2.33
Min	990	130	85	50	0
Max	420000	425600	175000	332300	320300
p25	15000	16200	7800	8000	990
p50	26000	28000	11200	12500	2724.5
p75	44000	46408	17800	20200	9000

Table 5. Descriptive statistics of income decomposition

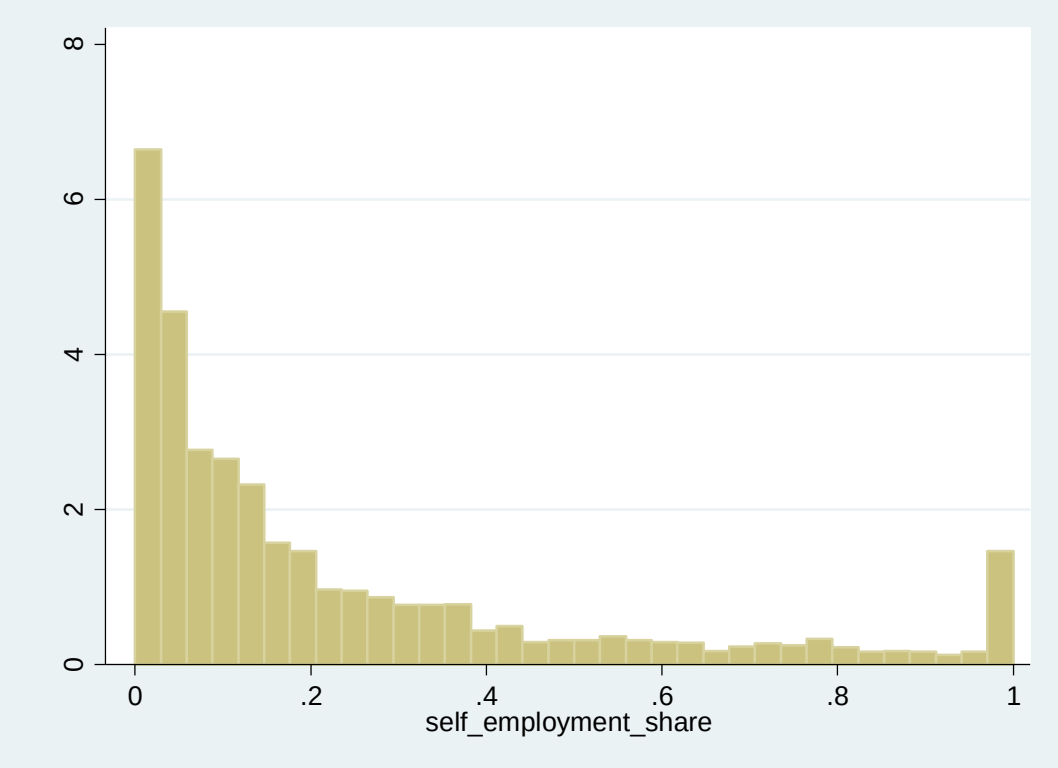
Now in our sample the amount of "zeros" from non wage income diminishes thanks to the fact that for some households total income has been increased by the inclusion of other income. On the other hand, the amount of zeros from self-employment income increases thanks to the fact that we have substituted the negative values by zero values because of mismatch between total income and its components.

According to the descriptive statistics table, the mean value of wage income is the greatest among all the mean values of various income sources: households who get wage report that they earn on the average 33700 rubles, whereas the mean for other income accounts for 13478 rubles, self employment income is on average 9405 rubles, and one must take into account the fact that this mean is only above those, who happen to have positive amount of income (near 45 % of the households). So on average, the income from self - employment is much less than the wage income. It must be found whether it is true or not and what part of household income from self - employment is hidden.

Coming to the variations of different components of household income we must notice that the assumptions proposed by Weber are fulfilled: self - employment part of income is much more *relative* variant. What does it mean? Although the variance of wage income is more than the variance of income from self - employment, the relative variance, or the variance related to the mean, is much greater for those who are self - employed (2.33 vs 0.89). On the other hand, the relative variance of other income, which is defined above and consists of the sum of pensions, scholarships and so on, is the least (0.71).

That is why we are able to perform the income decomposition according to the idea, proposed by Weber: we believe that the wage income is reported correctly and the self - employment income is under - reported, that is why we need to compute the multiplier to perform the income correction up to the true level of self - employment income.

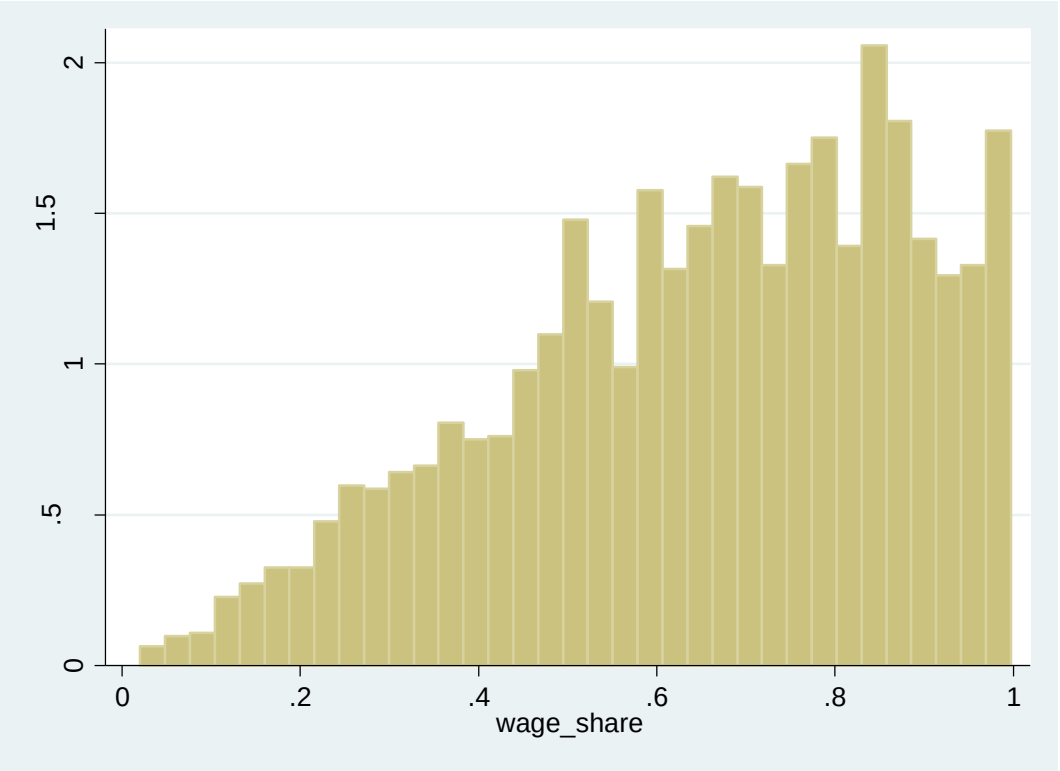
The next step of our model is to compute the part of total income which is associated to the self employment income. This will help to define the main part of household occupation. For those households who have self employment income one can perform a histogram of the income share distribution (condition if the share > 0) - Graph 1:



Graph 1. The density of self employment share of income distribution (self-employment income >0)

We can see that the distribution of the share of the self - employment income in the total income of household is skewed towards 1. It means that even if a household has some self - employment income, the majority of the households is still having it as a minor source of income.

The wage share distribution has an opposite skew in relation to the self employment income distribution. Have a look at the distribution of wage income share if it is either greater than 0 or less than 1 (Graph 2):



Graph 2. The density of wage share of income distribution (wage income share greater that 0 and less than 1)

The density of distribution increases with the two "steps" which occur at the share of 0.2 and 0.4.

Now we must determine the criteria, by which the household is self - employed. Following Webber, the criteria is such that the share of income from self-employment is greater than some threshold value. In our model this value is defined with the help of histogram of income share distribution. Let us say that this value is 0.2 (in the paper authors take the value of 0.25).

-> D_self_empl	0				
Variable	Obs	Mean	Std. Dev.	Min	Max
total_inc	1947	34532.18	27969.09	1000	276500
wage_inc	1947	20335.45	26283.81	0	243000
other_inc	1947	10769.81	10052.66	0	99000
self_employment_inc	1947	3426.922	8190.329	0	192000
-> D_self_empl	1				
Variable	Obs	Mean	Std. Dev.	Min	Max
total_inc	932	44003.35	50014.41	1500	425300
wage_inc	932	14866.87	24496.69	0	220000
other_inc	932	7232.321	9178.869	0	120000
self_employment_inc	932	21904.16	33352.1	600	320300

Table 6. Income decomposition descriptive statistics for the households with positive income from self employment, grouped by the share of self employment income in total income

Here there are the descriptive statistics of income decomposition for households with self-employment income (Table 6). Households with high self-employment income have mean higher total income and slightly less mean other income. We have seen earlier that self-employment income has higher coefficient of relative volatility than wage income. Surprisingly this is not true for our types of households with positive income from self employment. Wage income tends to have the same volatility as income from self - employment, but even higher relative volatility for those who are self - employed. Although self employment income is more volatile in second group, the relative variance is smaller. But total income for self employed part of sample is anyway more volatile (in absolute and relative terms). To sum up, the hypothesis of income variance holds but only for the absolute income variance (except for the variance in wage income).

To perform the income decomposition proposed by Webber one must be sure that the total income and the income earned by the self employed are log - normal. The hypothesis of log-normality of total income cannot be rejected: Kolmogorov - Smirnov test has a P-value of 14.7%. Whereas for the self - employment income the statistics has a P-value of 1.4%, log - normality also can be assumed.

Concluding with the descriptive statistics analyses we must say that the income of Russian households does not necessarily come with the British households income behavior. Especially it can be seen that wages may be as much volatile as self - employment income. This fact says that maybe some new estimation methods should be implemented to estimate the size of black economy.

Expenditure decomposition

As far as we know from the theory, the household expenditure can be decomposed into the expenditure on durable and nondurable goods. Let us try to perform this decomposition with respect to our database. In RLMS the items of expenditures are grouped by the categories. This simplifies the analyses and enables to join the items by the same expenditure purpose. The items we are interested in are food expenditures, clothes expenditures, service expenditures (all of them form the expenditures on nondurable goods). The durable goods expenditures are also needed and include expenditures on household appliances.

To start with let us describe the food expenditures in our sample. The households are interviewed on the amount of food products purchased, their prices and total expenditures. This information is obtained for the last seven days. Actually our aim is to analyze the information on the household expenditures on food products and to aggregate it. The question is how can we use the pricing information, since we are operating with the aggregated items. The pricing information is not used at the moment. The household expenditures on food during the month is just the sum of expenditures on all products bought normalized to the 30 days period. Here you can see the table of descriptive statistics of consumer expenditures on each item sorted in the order of the highest mean value (Table 7 in the application). The mean value and variance are calculated for all sample. You can see that there is just a part of items on which you have positive numbers of amount of money spent.

It is interesting to look at the amount of food products indicated as consumed by the consumers. "Consumption" means that people indicate positive amount of money spend. The rate of response to this part of interview is almost 100%. So the zero consumption is due to the fact that people indicate they do not consume those types of products.

We have generated the variable responsible for the amount of items, for which consumers have positive expenditures. The distribution of the variable can be seen in the table for the three parts of sample: 1) the people with the share of self - employment income from zero to 20 percent; 2) the people with the share of self - employment income above 20 percent; 3) the people with zero share of income from self - employment. So the table says that the median amount of items consumed is only 15 out of 57. Is that quite little? This is sufficient to build and estimate the model: this is quite sufficient for a budget survey, and the amount of average expenditures on consumption correspond to reality.

The main idea of the Table 8 (see application) is to find out whether the rate of responses of households on the questions depends on the occupation status: is it true that those who have self employment income and tend to deliberately reduce their incomes have the same rate of response on the questions concerning food consumption as those who have only wage income or pension? We see that cumulative distribution function of the number of questions answered by those who have zero percent of income from self - employment and 0-20 percent go slightly behind the function of the people with high share of self - employment income. It means that a typical self - employed is saying he is consuming 1 product item less than the man living on wage income. Of course, this is a minor difference. We must accept the hypothesis that self - employed have the same pattern of food consumption as the other income groups.

On the basis of the summation of these expenditures one may obtain the value of total expenditures of a household on food products. That is simply the amount of goods purchased by a household during a week. But as far as there is the detailed representation of food goods purchases, there is also a question of the amount of the expenditures on food outdoors during the

last seven days. More than that, there is the question: "how much the households spent both on food at home and outdoors during the last thirty days". There can be traced the difference between our computed value as the sum of household expenditures on different items and the calculated value from the database on food expenditures at home.

Variable	Obs	Obs > 0	Mean	Std. Dev.	Min	Max
food_total_30	6517	6000	8784	7341	0	73000
food_outdoors_30	6517	2366	1550	3839	0	77143
food_at_home_30	6517	5743	7233	7198	-53143	61785
food_exp_purchase_30	6517	6498	9518	8297	0	181847
food_total_30_comp	6517	6498	11069	9898	0	207561

Table 9. Food expenditure descriptive statistics

In table 9 there are represented the descriptive statistics on expenditures on food.

"food_total_30" - the variable in the interview, indicating how much the households spend on food during the last 30 days at home and outdoors;

"food_outdoors_30" - the variable in the interview, indicating how much the households spend on food outdoors during the last seven days (has been normalized to 30 days);

"food_at_home_30" = "food_total_30" - "food_outdoors_30";

"food_exp_purchase_30" - the sum of the households expenditures on the product items indicated in the interview (has been normalized to 30 days);

"food_total_30_comp" = "food_exp_purchase_30" + "food_outdoors_30".

First of all, the variables "food_total_30" and "food_outdoors_30" demonstrate that in the surveys the value of food expenditures has a "reasonable" upper bound of nearly 80000 rubles. But the computation of expenditures at home based on these two variables may be somehow misleading (the negative expenditures on food at home). On the other hand, there are a few households, indicating very high purchase of goods. This are the outliers and must be excluded from the sample the upper bound of the amount of household expenditures is defined by us at the value of 65000.

The use of the variable "food_at_home_30" can also be misleading thanks to the fact that the minimum value of this variable is negative (the negative values are quite a lot in the sample). So we suppose that the "food_outdoors_30" is the right variable, together with "food_exp_purchase_30". And on the basis total food expenditures are computed.

Let us have a look at the detailed descriptive statistics for the three parts of the sample: 1) the people with zero self-employment income; 2) the people with income share from self-employment from zero to 20%; 3) the people with income share from self - employment from 20% to 100% (Table 10).

self_employment_inc==0						
Variable	Obs	Obs > 0	Mean	Std. Dev.	Min	Max
food_total_30	3631	3278	9001	7455	0	50000
food_outdoors_30	3631	1333	1718	4351	0	77143
food_at_home_30	3631	3121	7284	7413	-53143	50000
food_exp_purchase_30	3631	3618	9462	7213	0	63570
food_total_30_comp	3631	3618	11180	9251	0	106590

self_employment_inc>0 & D_self_empl==0						
Variable	Obs	Obs > 0	Mean	Std. Dev.	Min	Max
food_total_30	1946	1843	8630	6732	0	65000
food_outdoors_30	1946	663	1284	2984	0	38571
food_at_home_30	1946	1782	7346	6504	-22571	61786
food_exp_purchase_30	1946	1943	9143	6882	0	54879
food_total_30_comp	1946	1943	10427	8080	0	62421

self_employment_inc>0 & D_self_empl==1						
Variable	Obs	Obs > 0	Mean	Std. Dev.	Min	Max
food_total_30	930	870	7970	7230	0	60000
food_outdoors_30	930	364	1402	3091	0	32143
food_at_home_30	930	832	6568	7062	-32143	60000
food_exp_purchase_30	930	927	9502	7624	0	62824
food_total_30_comp	930	927	10904	8868	0	63039

Table 10. Food expenditure descriptive statistics grouped by income groups

It should be noticed that self - employed do not demonstrate specific patterns in food consumption compared with the people with zero and low share of self - employment income. Maybe food consumption of self - employed is higher than such of those who have low share of self - employment income, but it is not evident that it is higher compared to the people with zero income from self - employment. That is why it is important to search for the other differences in expenditure patterns between the households.

Now we turn our attention to other components of household expenditures, among which there are the expenditures on clothes, services and durable goods. It is interesting to notice that people in the sample demonstrate extremely high rate of response. It means that almost all the households answer "yes" or "no", and just a few refuse to give an answer. But of course the answer "no" does not mean that the expenditures have not been made. So the value we are interested in is the average amount of household expenditures.

Now we come to the description of household expenditures items, apart from the food expenditures. According to the database, the expenditures are grouped the following way:

- Clothes expenditures (include clothes for adults and clothes for children, bought during the last 90 days);

- Durable goods expenditures (expenditures on the buying of TV, mobile phone, furniture, household appliances, automobile, motorcycle, garage, building materials, land/house, textbooks, bicycle), bought during the last 90 days. These expenditures are extremely volatile: 20789 is the mean value and the standard deviation is 125012. This is due to the fact that a few households indicate that they have very considerable expenditures on new car/automobile and so

on. That is why it is important to study more carefully those expenditure, namely we indicate the high expenditures on a category of durable goods if the household has spent on it more than 90000. The following categories are considered: automobile; motorcycle; garage; building materials; land/house. In the sample there are 160 households with high expenditures on one durable good, 6 with high expenditures on two durable goods, and 1 with high expenditures on three durable goods. The descriptive statistics for this part of sample (Table 11):

D_durable=1					
Variable	Obs	Mean	Std. Dev.	Min	Max
total_inc	167	73543	67463	0	425300
wage_inc	167	50068	43198	0	280000
other_inc	167	7498	12047	0	120000
self_employment_inc	167	15977	45549	0	320300

Table 11. Income descriptive statistics grouped by purchases of durable goods

For these observations a high level of mismatch between total expenditures and income is observed, so they should be excluded from the regression analyses in some cases.

This table is compared to the all sample descriptive for income decomposition (Table 12). As you can see, the self - employment income growth 4 times for this part of sample - it is the most considerable income component increase. But the extent of self - employment income under - reporting is still unclear.

All sample					
Variable	Obs	Mean	Std. Dev.	Min	Max
total_inc	6508	36082	33519	0	425600
wage_inc	6508	22720	28957	0	420000
other_inc	6508	9258	10072	0	175000
self_employment_inc	6508	4105	14891	0	320300

Table 12. Final descriptive statistics of household income decomposition

Apart from clothes and durable goods, the expenditure groups form the following categories:

- Service expenditures (includes transport, clothes repair, TV repair, house repair, auto repair, clothes wash, post, ritual service, mobile service, internet service, lower service, communal payment), bought during last 30 days;

- Other expenditures (children education, tourism, billets, washing, private things, cosmetics, adult courses, insurance, aliments, home tax) during last 30 days;

- Health care expenditures (medicaments, tooth care, 2 types of health care) during last 30 days.

In the application (Table 13) you can see the detailed expenditure summary statistics for the decomposition:

This is the mean value of expenditures by aggregated commodity groups (Table 14):

	all_sample	non_s.e.	0-20_s.e.	20-100_s.e.
clothes_exp (90 days)	7308	7457	6546	8310
durable_goods_exp (90 days)	20789	20599	17774	27854
service_exp (30 days)	7248	7499	6641	7529
health_care_exp (30 days)	2138	2127	2109	2247
other_exp (30 days)	2671	2765	2324	3024
Sum	21423	21743	19181	24855
food_exp_purchase_30	9518	9462	9143	9502
food_total_30_comp	11069	11180	10427	10904
Sum_2	32492	32923	29608	35759
total_inc	36082	35020	34514	43500

Table 14. Final descriptive statistics of household expenditure decomposition. Sum row is the sum of previous expenditure items. Sum 2 row is Sum+ food_total_30_comp. Food_exp_purchase_30 is part of total food expenditures. The total income below gives the idea how total expenditures and total income are related

The sample has been divided into three groups: non self - employed, people with low share of self employment income and people with high share of self - employment income (or self - employed). If we take all the people with any positive amount of self - employment income and compare them to all the sample, we can see that their expenditures do not seem to differ. On the other hand, there is high level of expenditure heterogeneity within the group of people who obtain income from self - employment. For those with high share of self - employment income (20 - 100%) the expenditures on all the types of categories are higher than both the all sample mean level (excluding food expenditures) and the people with low self - employment income share. On the other hand, the total income of self - employed is also considerably higher. So the question is whether the differences in the expenditure level may be attributed to the differences in reported income or there is some part unexplained by income variation, a part attributed to the black economy.

Econometric model estimation

Starting estimation the food equation let us once more provide the table of descriptive statistics for the expenditures (Table 15):

	all_sample	non_s.e.	0-20_s.e.	20-100_s.e.
clothes_exp (90 days)	7308	7457	6546	8310
durable_goods_exp (90 days)	20789	20599	17774	27854
service_exp (30 days)	7248	7499	6641	7529
health_care_exp (30 days)	2138	2127	2109	2247
other_exp (30 days)	2671	2765	2324	3024
Sum	21423	21743	19181	24855
food_exp_purchase_30	9518	9462	9143	9502
food_exp_purchase_new_30	9976	9907	9663	10905
food_outdoors_30	1543	1718	1283	1402
food_total_30_comp	11069	11180	10427	10904
food_total_new_30_comp	11519	11625	10946	12307
sum_2	32492	32923	29608	35759
sum_3	32942	33368	30127	37162
total_inc	36082	35020	34514	43500

Table 15. Final descriptive statistics of household expenditure decomposition accounting for home made goods consumption (sum 3)

This table indicates that food expenditures based on the computation with the inclusion of home production tend to be greater for those who are self - employed, although the difference is not striking. Self - employed tend to produce and consume more goods of their own production. But the estimation of food equation in this case may turn into the estimation of the part of shadow economy attributed to the income from agricultural production, because monetary expenditures on food for self - employed and other people do not differ. Sum_3 equals the sum of all expenditures with the inclusion of home made goods consumption, sum_2 is simply the expenditure sum without the consumption of home - made goods.

As you can see from the estimation of the food equation with identifying instruments, after determining the set of appropriate instruments, the Dummy - variable indicating the self - employed type is not significant at the level of 10% (see the estimation table). The reason for this may be such that there are the household characteristics influencing food expenditures along with the level of income, but the income matters up to a certain extent, if a household is too rich it does not increase the consumption of food. However, the quadratic term of income is not significant.

It means that the food equation estimation does not give information to estimate the black economy. But we shall try to estimate the equation for the other part of expenditures. The idea is to use the part which is the most correctly reported (that is why it is not the expenditures on durable goods), and at the same time depending on income but not the structure of income. That is why we cannot use service expenditures, which include transport expenditures and are linked to the workplace. Clothes expenditures perform as the most reasonable choice as measuring the household level of welfare.

VARIABLES	ln_food_total_new_30_comp
D_couple	0.124***
	(0.0196)
D_widow	-0.0397**
	(0.0201)
D_educ_higher	-0.0525***
	(0.0182)
D_white_collar	0.0809***
	(0.0184)
D_wish_other_job	0.117***
	(0.0176)
D_living_improvement	0.0662**
	(0.0268)
D_town	-0.114***
	(0.0184)
total_square	0.000883
	(0.000737)
living_square	0.000974
	(0.000945)
D_central_water	0.0869***
	(0.0275)
D_phone	-0.0708***
	(0.0174)
D_refrigerator	0.110***
	(0.0172)
D_washing_machine	0.0794***
	(0.0214)
D_plazma_TV	0.0372**
	(0.0183)
D_DVD	0.0585***
	(0.0173)
D_notebook	0.0844***
	(0.0192)
D_computer	0.0456**
	(0.0185)
D_bicycle	0.0797***
	(0.0203)
D_grown_smth	0.0655***
	(0.0180)
D_products_sold	0.178***
	(0.0565)
D_cabel_TV	0.0459**
	(0.0183)
D_self_empl	-0.0106
	(0.0226)
ln_Y_h	0.410***
	(0.0144)
Constant	4.419***
	(0.136)
Observations	6,192
R-squared	0.380

Standard errors inparentheses
*** p<0.01, ** p<0.05, * p<0.1

Here there is the estimation of food equation (Table 16). The income specification is linear, all significant variables identifying the households are kept in the model. The food expenditures are computed with the inclusion of consumption of home made goods.

The Dummy variable of self - employment is not significant on 10% significance level, therefore it does not provide any information on the estimation of the size of informal economy.

Table 16. Single food equation estimation

For the clothes expenditures the dependent variable is the expenditures on clothes for both children and adults made during the last 90 days. As far as we take the log of dependent variable, its scale is not important for the estimation.

The model of clothes expenditures enables to provide the significant coefficient of the increase in the level of clothes expenditures for the self-employed compared to all the sample. To perform the clothes expenditure estimation we proceed the following way:

1) estimate the model on the full sample and call it "general";
2) identify the set of significant variables identifying the household for this sample;
3) divide the sample into the subsamples according with key types of household;
4) hold the same identifying variables in the model to obtain comparable results;
5) for each model obtain the value and significance of black economy dummy variable and the marginal propensity to consume.

Then for each subsample the black economy coefficient must be estimated. To make this we need:

1) To perform the supplementary regression for income, using exactly the same set of identifying variables. The aim is to estimate the residual error term.
2) Obtain the residual income variance estimate from STATA output as root MSE. The variance is obtained separately for the subsample of self - employed and for the subsample of employees. Along with the theory, the residual error variance for self - employed is greater than for non self-employed for all types of the models.
3) To compute the mean under - reporting parameter and the lower and upper bounds of it as determined on the base of analytical expression obtained by Pissarides and Webber.
4) With the help of these bounds we know the parameter by which the income of self employed part of households needs to be multiplied to obtain the corrected income of self - employed.
5) The mean income of self - employed and employees for each part of sample is defined and the share of self - employed households in the population.
6) The income of the self - employed is corrected with respect to this multiplier and the mean income of all sample is computed.

The bounds for the parameter of income under - estimation are computed in the tables below.

	(1)	(2)	(3)
SAMPLE CONDITION		D_white_collar=1	D_white_collar=0
SAMPLE CONDITION 2			D_active=1
THE MODEL	General	white_collar	blue_collar_active
VARIABLES			
popul	-0.0166***	-0.0252***	0.0058
	(0.005)	(0.007)	(0.008)
D_number_of_children	0.0414**	0.0575***	0.0408
	(0.016)	(0.022)	(0.026)
D_wish_other_job	0.1092***	0.0783**	0.0805
	(0.032)	(0.040)	(0.054)
D_child_study_pay	0.1740***	0.1213***	0.2632***
	(0.038)	(0.043)	(0.072)
D_city	-0.0831**	-0.0357	-0.1110*
	(0.037)	(0.047)	(0.062)
D_rent_house	0.1542***	0.0883	0.1827*
	(0.057)	(0.072)	(0.094)
living_square	0.0028***	0.0027**	0.0031*
	(0.001)	(0.001)	(0.002)
D_phone	-0.1770***	-0.0641	-0.2745***
	(0.034)	(0.043)	(0.056)
D_washing_machine	0.1326***	0.0652	0.1177*
	(0.040)	(0.056)	(0.063)
D_DVD	0.1402***	0.0417	0.1851***
	(0.032)	(0.039)	(0.053)
D_digital_camera	0.2231***	0.2149***	0.2073***
	(0.036)	(0.043)	(0.062)
D_notebook	0.2741***	0.1691***	0.3520***
	(0.035)	(0.043)	(0.060)
D_computer	0.3163***	0.1622***	0.4268***
	(0.034)	(0.043)	(0.057)
D_bicycle	0.1407***	0.1766***	0.0695
	(0.036)	(0.044)	(0.065)
D_grown_smth	-0.1515***	-0.1586***	-0.1160**
	(0.034)	(0.042)	(0.058)
D_self_empl	0.0839**	0.1090*	0.0355
	(0.042)	(0.058)	(0.061)
ln_Y_h	0.5403***	0.4990***	0.4666***
	(0.026)	(0.035)	(0.044)
Constant	2.4283***	3.1065***	3.1133***
	(0.250)	(0.353)	(0.406)
Observations	4,223	2,317	1,663
R-squared	0.3154	0.2072	0.3082
Adj. R-squared	0.3126	0.2013	0.3011
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

Table 17. 1rst part of the clothes expenditures model estimation (full sample base)

THE MODEL		General	white_collar	blue_collar_active
σ_{Yse}	$root\ MSE\ s.e.$	0.676	0.673	0.642
σ_{Yee}	$root\ MSE\ e.e.$	0.556	0.515	0.549
${\sigma_{Yse}}^2$	$MSE\ s.e.$	0.457	0.452	0.412
${\sigma_{Yee}}^2$	$MSE\ e.e.$	0.309	0.265	0.302
γ_j	$inc.underreport$	0.0839	0.109	0.0355
β_j	mpc	0.5403	0.499	0.4666
γ_j / β_j	$mean\ ln\bar{k}$	0.155	0.218	0.076
$\frac{\gamma_j}{\beta_j} - \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ lower\ bound$	0.082	0.125	0.021
$\frac{\gamma_j}{\beta_j} + \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ upper\ bound$	0.229	0.312	0.131
	$mean\ \bar{k}$	1.168	1.244	1.079
	$\bar{k}\ lower\ bound$	1.085	1.133	1.021
	$\bar{k}\ upper\ bound$	1.257	1.366	1.140
Sample size		6508	2982	2649
Number of self_employed		930	355	575
Percentage of self_employed		14.29%	11.90%	21.71%
Mean total income		36082	47591	30237
Mean total income of self employed		43500	59691	33503
Mean total income of others		34845	45956	29332
Lower bound mean total income of s. e.		47195	67623	34205
Upper bound mean total income of s. e.		54697	81555	38209
Lower bound mean total income		36610	48535	30389
Upper bound mean total income		37682	50194	31258
Lower bound black economy estimate		1.46%	1.98%	0.50%
Upper bound black economy estimate		4.43%	5.47%	3.38%

Table 18. 1rst part of the black economy coefficients estimate (full sample base)

In the tables 17 and 19; 18 and 20 you can see the results of the **first part of clothes expenditure model estimation** and the black economy coefficient estimates based on it. The base sample includes all the households (the estimation on those with positive expenditures on clothes). The second and third subsample perform an estimation of clothes expenditure functions based on the households which members are working as white collars and blue collars respectively (the third is an active household but not a white collar). The fourth one is devoted to the head of the household being working as a white collar. The fifth is computed on the subsample of the people which are married (form a couple). The sixth subsample is the people which are single or divorced. At last, the seventh subsample is single and blue collar occupied. At the bottom of the tables of estimates you see the number of observations and the model R - squared.

	(4)	(5)	(6)	(7)
SAMPLE CONDITION	D_white_ collar=1	D_ couple=1	D_ couple=0	D_couple=0
SAMPLE CONDITION 2	D_male=1			D_white_collar=0
THE MODEL	white_collar _male	couple	non_couple	non_couple_ blue_collar
VARIABLES				
popul	-0.0287***	-0.0275***	-0.0047	0.0118
	(0.008)	(0.007)	(0.008)	(0.011)
D_number_of_children	0.0491**	0.0647***	0.0176	0.0326
	(0.025)	(0.022)	(0.025)	(0.034)
D_wish_other_job	0.0940**	0.0992**	0.1184**	0.0921
	(0.046)	(0.040)	(0.053)	(0.074)
D_child_study_pay	0.0903*	0.1852***	0.1563**	0.3311***
	(0.049)	(0.044)	(0.074)	(0.113)
D_city	-0.0454	-0.0320	-0.1498**	-0.0814
	(0.055)	(0.048)	(0.059)	(0.083)
D_rent_house	0.0117	0.1214	0.1689*	0.1702
	(0.081)	(0.077)	(0.087)	(0.117)
living_square	0.0021*	0.0034***	0.0024	0.0006
	(0.001)	(0.001)	(0.002)	(0.003)
D_phone	-0.0047	-0.1153***	-0.2485***	-0.2780***
	(0.049)	(0.043)	(0.055)	(0.077)
D_washing_machine	0.0060	0.0864	0.1751***	0.1788**
	(0.064)	(0.058)	(0.058)	(0.079)
D_DVD	0.0389	0.1417***	0.1349***	0.1679**
	(0.046)	(0.040)	(0.052)	(0.073)
D_digital_camera	0.2279***	0.1730***	0.3039***	0.3109***
	(0.051)	(0.045)	(0.059)	(0.084)
D_notebook	0.2062***	0.2290***	0.3338***	0.4330***
	(0.050)	(0.044)	(0.058)	(0.083)
D_computer	0.1551***	0.3285***	0.3050***	0.4136***
	(0.049)	(0.043)	(0.056)	(0.078)
D_bicycle	0.1459***	0.1653***	0.0839	0.0104
	(0.051)	(0.044)	(0.066)	(0.097)
D_grown_smth	-0.1211**	-0.2052***	-0.0749	0.0198
	(0.048)	(0.042)	(0.056)	(0.084)
D_self_empl	0.0439	0.0193	0.1475**	0.1587**
	(0.067)	(0.056)	(0.064)	(0.079)
ln_Y_h	0.4528***	0.5338***	0.5515***	0.4583***
	(0.042)	(0.037)	(0.041)	(0.056)
Constant	3.6391***	2.4925***	2.3276***	3.1189***
	(0.417)	(0.365)	(0.381)	(0.518)
Observations	1,689	2,482	1,741	920
R-squared	0.1909	0.2674	0.3228	0.3277
Adj. R-squared	0.1827	0.2623	0.3161	0.3150
Standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

Table 19. 2nd part of the clothes expenditures model estimation (full sample base)

THE MODEL		white_collar _male (4)	couple (5)	non_ couple (6)	non_couple_ blue_collar (7)
σ_{Yse}	$root\ MSE\ s.e.$	0.670	0.647	0.653	0.622
σ_{Yee}	$root\ MSE\ e.e.$	0.505	0.499	0.562	0.549
${\sigma_{Yse}}^2$	$MSE\ s.e.$	0.449	0.419	0.426	0.387
${\sigma_{Yee}}^2$	$MSE\ e.e.$	0.255	0.249	0.316	0.302
γ_j	$inc.underreport$	0.0439	0.0193	0.1475	0.1587
β_j	mpc	0.4528	0.5338	0.5515	0.4583
γ_j / β_j	$mean\ ln\bar{k}$	0.097	0.036	0.267	0.346
$\frac{\gamma_j}{\beta_j} - \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ lower\ bound$	0.000	-0.049	0.212	0.304
$\frac{\gamma_j}{\beta_j} + \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ upper\ bound$	0.194	0.121	0.323	0.389
	$mean\ \bar{k}$	1.102	1.037	1.307	1.414
	$\bar{k}\ lower\ bound$	1.000	0.953	1.237	1.355
	$\bar{k}\ upper\ bound$	1.214	1.128	1.381	1.475
Sample size		2168	3435	3071	1585
Number of self_employed		256	449	481	374
Percentage of self_employed		11.81%	13.07%	15.66%	23.60%
Mean total income		48785	44600	26570	24525
Mean total income of self employed		60074	55736	36976	26953
Mean total income of others		47274	42925	24637	23775
Lower bound mean total income of s. e.		60088	53094	45722	36518
Upper bound mean total income of s. e.		72911	62897	51052	39763
Lower bound mean total income		48787	44255	27940	26782
Upper bound mean total income		50301	45536	28775	27548
Lower bound black economy estimate		0.00%	-0.77%	5.16%	9.20%
Upper bound black economy estimate		3.11%	2.10%	8.30%	12.32%

Table 20. 2nd part of the black economy coefficients estimate (full sample base)

The common idea is that for the general model for all sample we indicate the set of identifying variables, which are kept for the model. In the first model they are: population, number of children; individual in the household characteristics - wish for other job, be able to pay for a study of a child; living conditions - a households lives in the city, rents house, the living square; possession of durable goods - phone, washing machine, DVD, digital camera, notebook, computer, bicycle; the growth of home - made products. The variables of the model are dummy of self - employment, logarithm of total income.

Actually we do not consider the sign interpretation of certain coefficients, since they are included to avoid the bias in the shadow economy coefficient and log income coefficient. Also the square of log income is not included because of multicollinearity and the significance loss of income coefficient.

Noticeably, for each subsample in the model the auxiliary regressions for income give the same ratio between the residual income variance for self - employed and employees, which comes along with the hypothesis of Pissarides and Webber model. But the residual variance of income for self - employed is not much than 2 times higher - a small difference to be compared with Pissarides and Webber.

Self employed dummy is significant at the 5% level for the first model total sample, significant at the 10% level for the white collar subsample and significant at the 5% level for the subsample of non - couple: single individuals. Unfortunately, for head white collar and single white collar we could not obtain the significant estimate of income under report coefficient. So the model for subsample of single blue collar has been estimated, for which the black economy coefficient was the largest and also significant (at 5% level).

The income under-report coefficients vary: from 8.5% to 25.7% for all the sample; from 13.3% to 36.6% for white collar; from 23.7% to 38.1% for single people; from 35.5% to 47.5% for single blue collar. Whereas the black economy estimate varies from 1.46% to 4.43% for all the sample; from 1.98% to 5.47% for the white collar subsample; from 5.16% to 8.30% for single people; from 9.20% to 12.32% for single blue collar. The estimates of income - under reporting coefficient are corresponding to the black economy estimates.

The factors which determine the size of informal economy include the following ones: 1) the value of income under - reporting coefficient (the higher the value, the more income is under reported) 2) the share in the sample of self employed relative to the total amount of households (the more the share, the more the share of people attributed to "black economy") 3) the mean income of self - employed household in comparison with the mean income of the whole sample (the more self - employed earn, the more their income is corrected due to the under - reporting coefficient).

	(1)	(2)	(3)
SAMPLE CONDITION		D_white_collar=1	D_white_collar=0
SAMPLE CONDITION 2			D_active=1
THE MODEL	General	white_collar	blue_collar_active
VARIABLES			
D_educ_sec_special	-0.1089**	-0.0674	-0.1577**
	(0.046)	(0.058)	(0.072)
D_number_of_children	0.0611***	0.0946***	0.0305
	(0.022)	(0.030)	(0.032)
D_child_study_pay	0.2884***	0.2303***	0.3648***
	(0.057)	(0.065)	(0.103)
D_city	-0.1250**	-0.0847	-0.1461*
	(0.049)	(0.064)	(0.075)
D_rent_house	0.2238**	0.0566	0.3217**
	(0.087)	(0.110)	(0.136)
D_phone	-0.2228***	-0.0665	-0.3281***
	(0.049)	(0.064)	(0.076)
D_DVD	0.1679***	0.0666	0.2207***
	(0.047)	(0.061)	(0.072)
D_digital_camera	0.2656***	0.2818***	0.2231**
	(0.054)	(0.067)	(0.087)
D_notebook	0.3452***	0.1985***	0.4488***
	(0.053)	(0.066)	(0.084)
D_computer	0.3496***	0.1719***	0.4581***
	(0.051)	(0.065)	(0.080)
D_self_empl	0.0961**	0.1303**	0.0599
	(0.048)	(0.063)	(0.072)
ln_Y_h	0.5295***	0.4468***	0.5473***
	(0.035)	(0.049)	(0.053)
Constant	2.6628***	3.6485***	2.4813***
	(0.339)	(0.496)	(0.499)
Observations	1,955	963	992
R-squared	0.3230	0.2182	0.3251
Adj. R-squared	0.3189	0.2083	0.3168
Standard errors in parentheses			
*** p<0.01, ** p<0.05, * p<0.1			

Table 21. 1rst part of the clothes expenditures model estimation (only self_employed sample base)

THE MODEL		General	white_collar	blue_collar_active
σ_{Yse}	$root\ MSE\ s.e.$	0.720	0.699	0.701
σ_{Yee}	$root\ MSE\ e.e.$	0.585	0.529	0.571
${\sigma_{Yse}}^2$	$MSE\ s.e.$	0.519	0.488	0.492
${\sigma_{Yee}}^2$	$MSE\ e.e.$	0.342	0.280	0.327
γ_j	$inc.underreport$	0.0961	0.1303	0.0599
β_j	mpc	0.5295	0.4468	0.5473
γ_j / β_j	$mean\ ln\bar{k}$	0.181	0.292	0.109
$\frac{\gamma_j}{\beta_j} - \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ lower\ bound$	0.093	0.188	0.027
$\frac{\gamma_j}{\beta_j} + \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ upper\ bound$	0.270	0.396	0.192
	$mean\ \bar{k}$	1.199	1.339	1.116
	$\bar{k}\ lower\ bound$	1.098	1.206	1.027
	$\bar{k}\ upper\ bound$	1.310	1.485	1.212
Sample size		6508	2982	2649
Number of self_employed		930	355	575
Percentage of self_employed		14.29%	11.90%	21.71%
Mean total income		36082	47591	30237
Mean total income of self employed		43500	59691	33503
Mean total income of others		34845	45956	29332
Lower bound mean total income of s. e.		47751	72006	34416
Upper bound mean total income of s. e.		56969	88665	40595
Lower bound mean total income		36689	49057	30435
Upper bound mean total income		38007	51040	31776
Lower bound black economy estimate		1.68%	3.08%	0.66%
Upper bound black economy estimate		5.33%	7.25%	5.09%

Table 22. 1rst part of the black economy coefficients estimate (only self_employed sample base)

In the tables 21 and 23; 22 and 24 you can see the results of the **second part of clothes expenditure model estimation.** The difference is that as the base sample we use the sample of people with positive amount of self - employed income (in this sample there are people with low share of self - employment income and high share of self employed income - which are classified as "pure" self - employed). The number of observations to compute the shadow economy size is kept the same, but the estimation is produced on a smaller sample covering all the initial range of observations of people with positive income from self - employment.

The principle of the dividing the sample into subsamples is exactly the same as for the general model in the first part of expenditure model estimation. The set of identifying instruments is such to fit the narrower sample of individuals. The instruments are: individual member household characteristics - the education secondary special of one of the households, number of children in the household, possibility to pay for the study of a child; living conditions - a household lives in a city, rents a house; the durable goods in the household - phone, DVD,

digital camera, notebook, computer. The variables of the model are as before Dummy self-employment and log of total income.

Self - employment dummy is significant at the 5% level for the general sample and white collar sample, and on the 10% level for the non couple sample and non couple blue collar sample.

	(4)	(5)	(6)	(7)
SAMPLE CONDITION	D_white_collar=1	D_couple=1	D_couple=0	D_couple=0
SAMPLE CONDITION 2	D_male=1			D_white_collar=0
THE MODEL	white_collar_male	couple	non_couple	non_couple_blue_collar
VARIABLES				
D_educ_sec_special	-0.0983	-0.1367**	-0.0413	-0.0808
	(0.067)	(0.058)	(0.076)	(0.094)
D_number_of_children	0.0961***	0.1020***	0.0352	0.0278
	(0.034)	(0.030)	(0.034)	(0.040)
D_child_study_pay	0.2135***	0.2766***	0.3147***	0.4351***
	(0.073)	(0.066)	(0.110)	(0.151)
D_city	-0.0931	-0.1760***	-0.0926	-0.0539
	(0.073)	(0.065)	(0.077)	(0.096)
D_rent_house	-0.0379	0.1493	0.2855**	0.2324
	(0.119)	(0.115)	(0.133)	(0.163)
D_phone	-0.0836	-0.1454**	-0.2942***	-0.2721***
	(0.072)	(0.064)	(0.078)	(0.096)
D_DVD	0.0699	0.0942	0.2665***	0.2545***
	(0.070)	(0.061)	(0.074)	(0.093)
D_digital_camera	0.3115***	0.2089***	0.3560***	0.3308***
	(0.076)	(0.069)	(0.088)	(0.112)
D_notebook	0.2509***	0.2287***	0.4631***	0.5828***
	(0.075)	(0.068)	(0.086)	(0.109)
D_computer	0.1862**	0.4074***	0.2983***	0.4331***
	(0.074)	(0.065)	(0.083)	(0.104)
D_self_empl	0.0581	0.0359	0.1424*	0.1568*
	(0.073)	(0.063)	(0.074)	(0.089)
ln_Y_h	0.4025***	0.5638***	0.5119***	0.5572***
	(0.057)	(0.051)	(0.054)	(0.066)
Constant	4.1020***	2.3102***	2.7722***	2.2157***
	(0.579)	(0.506)	(0.500)	(0.605)
Observations	725	1,116	839	580
R-squared	0.2214	0.2807	0.3330	0.3760
Adj. R-squared	0.2082	0.2729	0.3233	0.3627
Standard errors in parentheses				
*** p<0.01, ** p<0.05, * p<0.1				

Table 23. 2nd part of the clothes expenditures model estimation (only self_employed sample base)

THE MODEL		white_collar_male	couple	non_couple	non_couple_blue_collar
σ_{Yse}	$root\ MSE\ s.e.$	0.691	0.665	0.719	0.690
σ_{Yee}	$root\ MSE\ e.e.$	0.511	0.521	0.584	0.551
${\sigma_{Yse}}^2$	$MSE\ s.e.$	0.477	0.442	0.517	0.476
${\sigma_{Yee}}^2$	$MSE\ e.e.$	0.261	0.271	0.341	0.304
γ_j	$inc.underreport$	0.0581	0.0359	0.1424	0.1568
β_j	mpc	0.4025	0.5638	0.5119	0.5572
γ_j / β_j	$mean\ ln\bar{k}$	0.144	0.064	0.278	0.281
$\frac{\gamma_j}{\beta_j} - \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ lower\ bound$	0.036	-0.022	0.190	0.195
$\frac{\gamma_j}{\beta_j} + \frac{1}{2}({\sigma_{Yse}}^2 - {\sigma_{Yee}}^2)$	$ln\bar{k}\ upper\ bound$	0.253	0.149	0.366	0.368
	$mean\ \bar{k}$	1.155	1.066	1.321	1.325
	$\bar{k}\ lower\ bound$	1.037	0.979	1.210	1.216
	$\bar{k}\ upper\ bound$	1.287	1.161	1.442	1.444
Sample size		2168	3435	3071	1585
Number of self_employed		256	449	481	374
Percentage of self_employed		11.81%	13.07%	15.66%	23.60%
Mean total income		48785	44600	26570	24525
Mean total income of self employed		60074	55736	36976	26953
Mean total income of others		47274	42925	24637	23775
Lower bound mean total income of s. e.		62284	54545	44726	32765
Upper bound mean total income of s. e.		77335	64688	53321	38926
Lower bound mean total income		49046	44444	27784	25896
Upper bound mean total income		50823	45770	29130	27350
Lower bound black economy estimate		0.53%	-0.35%	4.57%	5.59%
Upper bound black economy estimate		4.18%	2.62%	9.64%	11.52%

Table 24. 2nd part of the black economy coefficients estimate (only self_employed sample base)

Self - employed dummy is significant as before for exactly the same subsamples. They are the total sample, the subsample of white collar households, non couple households and non couple blue collar households. The value of the coefficients is nearly the same, but for some models the coefficients are larger. It means that we have stable ratios between the shadow economy estimates on a sample of self - employed people and the whole household database.

The value of income underreport by self - employed is from 9.8% to 31% for general sample; from 20.6% from 48.5% for the white collar sample; from 21% to 44.2% for non couples; from 21.6% to 44.4% for non couple blue collar. The black economy estimates vary from 1.68% to 5.33% for all sample; from 3.08% to 7.25% for the sample of white collar workers; from 4.57% to 9.64% for the sample of single people; and from 5.59% to 11.52% for the blue collar single.

The model parameters estimated on a smaller sample of people with positive income from self - employment indicate nearly the same ratios of income underreport and black economy coefficients. This indicates that the underreport patterns are stable to the estimation method proposed, and individuals with different occupation and family status share the different patterns for participation in the informal economy activities. Which is quite logical, although the surprising fact is that blue collar workers do not participate in informal economy when they live together, but start participating when they are single. For the white collar workers, the opposite is true, that single do not participate in the informal economy. Although overall informal economy participation ratio is greater for those who do not live in the couple.

The next step of our research is to consider not the different coefficients estimation results but the different sample for the black economy. According to the model of Pissarides and Webber, 1989, the individual has two parts of income: wage income and self-employed income. But this decomposition is true only for the 44% of households in RLMS sample. So this fact implies that this model is more likely to be implemented to the subsample of self-employed only. That is why we try to represent the results of black economy estimates computed solely on the subsample of self - employed (not only the model coefficients). To sum up we represent these results in two tables, indicating all the possible four cases. The coefficients are estimated either on the full sample (k1) or the self - employed sample (k2). The incomes and the share of black economy are estimated either on full sample (Part 1. Sample size 1) (Table 25) or on the self - employed sample only (Part 2. Sample size 2) (Table 26).

		(1)	(2)	(6)	(7)
		general	white_ collar	non_ couple	non_couple_ blue_collar
Part1. Sample size 1					
k lower 1		1.085	1.133	1.237	1.355
k upper 1		1.257	1.366	1.381	1.475
k lower 2		1.098	1.206	1.210	1.216
k upper 2		1.310	1.485	1.442	1.444
Lower bound black economy estimate k1		1.46%	1.98%	4.47%	9.20%
Upper bound black economy estimate k1		4.43%	5.47%	7.20%	12.32%
Lower bound black economy estimate k2		1.68%	3.08%	3.96%	5.59%
Upper bound black economy estimate k2		5.33%	7.25%	8.36%	11.52%

Table 25. The informal economy estimates computed for the whole sample

The estimation results obtained for the sample size 1 represent the same ones obtained in the tables 17-24 but in the form convenient for the comparison.

The results for the sample size 2 are the results when the subsample of self -employed only is considered as the total sampler of households. Due to the fact that the sample is cut to the size of almost 44% relative to the initial sample, Thus we obtain the more "realistic" estimates of the size of black economy - up to 10 - 12% of the total income. For the single blue collar workers the value is up to 18 - 20% of total income. Non - couples and white collar workers also indicate a greater income under - reporting ratio. For the estimates got with the help of second coefficient (k2) the distance between the bounds is larger, indicating the fact that the difference between the residual variance of self - employed income and employee income is greater.

		(1)	(2)	(6)	(7)
		general	white_ collar	non_ couple	non_couple_ blue_collar
Part2. Sample size 2					
k lower 1		1.085	1.133	1.237	1.441
k upper 1		1.257	1.366	1.381	1.605
k lower 2		1.098	1.206	1.210	1.216
k upper 2		1.310	1.485	1.442	1.444
Lower bound black economy estimate k1		3.19%	4.68%	9.63%	14.87%
Upper bound black economy estimate k1		9.67%	12.90%	15.50%	19.91%
Lower bound black economy estimate k2		3.67%	7.27%	8.53%	9.03%
Upper bound black economy estimate k2		11.63%	17.10%	17.99%	18.61%

Table 26. The informal economy estimates computed for the self-employed sample only

The obtained results have a drawback that is they do not account for wage income under report. To make the wage income be underreported, we shall try to estimate the system of consumer budget share equations, since the single equation approach is too simplified for this purposes. The system of consumer demand equations allows for the shift between demand on the different consumption goods depending on the income and individual occupation group. Also it uses the income decomposition but not the arbitrarily status of a household, whether it is self - employed or not. Thus it provides more accurate estimate of informal economy.

The main object of our interest is the under-reporting parameters of wage income and self -employment income. The model provides separate estimates of both these parameters. On the first step of estimation we do not account for preference heterogeneity (the constant term does not depend on the type of a household). The system is estimated on the expenditures on non - durable goods consumption, as far as these expenditures are less volatile and more correctly indicated by the households. The expenditure categories are: food expenditures, clothes expenditures, service expenditures, health care expenditures, other expenditures. Other expenditures are more likely to indicate the luxury goods expenditures, since include the expenditures on tourism.

The model is written as follows:

$$w_{ih} = \alpha_i + \sum_j \alpha_{ij} z_{hj} + \beta_i [lnY_h + ln(\theta_0 y_h^w + \theta_1 y_h^s)] + v_{ih}$$

The constant term is the same for each expenditure category, the model does not include quadratic income term equation; $\theta_0 \geq 1$ and $\theta_1 \geq 1$. The system is estimated through nonlinear seemingly unrelated equations method (nlsur), by feasible generalized nonlinear least squares (FGNLS). Nonlinearity in the coefficients arises from the θ parameters staying inside the log of income expression for the consumption function. The nonlinear estimation approaches allow for the separate estimation of marginal propensity to consume and both θ.

The sample used for the estimation of the model is only the sample of individuals with positive income from self - employment. For technical aspects of the model the sample is restricted to the households which have positive expenditures on clothes and food. The identifying variables characterizing households are introduced in the model.

The estimation procedure is quite tricky and demands for the appropriate set of initial conditions.

For the model the following identifying variables have been chosen, to manage to fit consumption patterns of households with different background. The choice of variables is based

on the previous single clothes equation estimation on the two samples: the large sample and the smaller subsample of households with self - employed income. The variables were tested on significance and were kept to be significant for most of the equations. They are: number of children in the household (D_number_of_children); the possibility to pay for a child study (D_child_study_pay); white collar worker in the family (D_white_collar); the household living in the town (D_town); the household owing notebook (D_notebook). The estimation is carried through all the sample of people with positive income from self-employment, regardless of the family status and the occupation category (these variables are included in the set of identifying variables). Therefore the under - reporting parameters do not vary from one group of people to another. That is made to increase the computing power of the model with the limited number of observations.

As has been mentioned, the model is estimated without the preference heterogeneity parameters. Although the inclusion in each equation the variable D_self_employment can reflect the preference heterogeneity, the advantage of the simultaneous equation estimation method, this variable is not significant. Therefore, the preference heterogeneity can be reflected through the use of household identifying variables.

Let us specify the system of equations:

$$
\begin{aligned}
w_{food,h} = {} & \{a0\} + \{a1\} * D(number\ of\ children) + \{a2\} * D(child\ study\ pay) + \{a3\} \\
& * D(white\ collar) + \{a4\} * D(town) + \{a5\} * D(notebook) + \{a6\} * (lnY_h \\
& + ln(\theta_0 y_h^w + \theta_1 y_h^s)) + v_{clothes,h} \\
w_{clothes,h} = {} & \{b0\} + \{b1\} * D(number\ of\ children) + \{b2\} * D(child\ study\ pay) + \{b3\} \\
& * D(white\ collar) + \{b4\} * D(town) + \{b5\} * D(notebook) + \{b6\} * (lnY_h \\
& + ln(\theta_0 y_h^w + \theta_1 y_h^s)) + v_{clothes,h} \\
w_{service,h} = {} & \{d0\} + \{d1\} * D(number\ of\ children) + \{d2\} * D(child\ study\ pay) + \{d3\} \\
& * D(white\ collar) + \{d4\} * D(town) + \{d5\} * D(notebook) + \{d6\} * (lnY_h \\
& + ln(\theta_0 y_h^w + \theta_1 y_h^s)) + v_{service,h} \\
w_{health\ care,h} = {} & \{e0\} + \{e1\} * D(number\ of\ children) + \{e2\} * D(child\ study\ pay) + \{e3\} \\
& * D(white\ collar) + \{e4\} * D(town) + \{e5\} * D(notebook) + \{e6\} * (lnY_h \\
& + ln(\theta_0 y_h^w + \theta_1 y_h^s)) + v_{health\ care,h} \\
w_{other\ exp,h} = {} & \{f0\} + \{f1\} * D(number\ of\ children) + \{f2\} * D(child\ study\ pay) + \{f3\} \\
& * D(white\ collar) + \{f4\} * D(town) + \{f5\} * D(notebook) + \{f6\} * (lnY_h \\
& + ln(\theta_0 y_h^w + \theta_1 y_h^s)) + v_{other\ exp,h}
\end{aligned}
$$

For each equation the coefficients are special, the common terms are only θ_0 and θ_1. The letter "c" is reserved for durable goods expenditures.

In the table of the system of nonlinear equations estimates (Table 27) there are the estimates of all the five equations (durable goods expenditures are not included in the model). Actually the coefficients of interest are the parameters θ and the $6th$ coefficient in the each equation, which indicated how the budget share is related to the income. It says what the expenditure category is the luxury goods category and what are necessity goods for the consumer. The luxury goods expenditures share increases with the income of a household (income elasticity is greater than 1), whereas the necessity goods decreases (income elasticity is less than 1).

The quadratic income term has not been included to simplify the estimation. As have been mentioned before, the parameters estimates are the same for all the subsamples. But as far as we know the relation between income underreport for different groups of population from the single clothes equation, this information can be used to compute the income underreporting for these groups.

Food		clothes		service		health care		other exp	
a0	0.7372***	b0	0.0321*	d0	0.2639***	e0	0.1467***	f0	-0.0604***
	(0.037)		(0.017)		(0.028)		(0.023)		(0.016)
a1	0.0214***	b1	0.0041	d1	-0.0223***	e1	-0.0012	f1	-0.0061***
	(0.004)		(0.003)		(0.003)		(0.002)		(0.002)
a2	-0.0238**	b2	0.0181***	d2	-0.0162**	e2	-0.0069	f2	0.0068
	(0.010)		(0.006)		(0.008)		(0.005)		(0.005)
a3	0.0234**	b3	-0.0074	d3	-0.0117	e3	-0.0147***	f3	0.0006
	(0.009)		(0.005)		(0.007)		(0.005)		(0.005)
a4	-0.0486***	b4	-0.0001	d4	0.0320***	e4	0.0079	f4	0.0024
	(0.009)		(0.004)		(0.007)		(0.005)		(0.004)
a5	-0.0503***	b5	0.0047	d5	0.0195***	e5	-0.0024	f5	0.0112**
	(0.009)		(0.005)		(0.007)		(0.005)		(0.005)
a6	-0.0313***	b6	0.0065***	d6	-0.0004	e6	-0.0071***	f6	0.0135***
	(0.004)		(0.002)		(0.003)		(0.002)		(0.002)
teta0	1.2242***								
	(0.273)								
teta1	1.4117								
	(0.000)								
Obs	1952								

Table 27. The consumer demand approach FGNLS estimates

The income terms indicate that the necessities are food and health care. Whereas the luxury goods are clothes and other goods. Services are neither of them. This explains the idea why the estimation of the food equation has given us wrong results. The demand for food increases at the slower rate than the demand for other goods, thus the food share drops as the income grows. For clothes the situation is opposite, when the income grows, the budget share is also growing. For both cases the size of informal economy may be computed, also the growing income share makes the estimation easier. We have seen that obtaining the positive estimates of informal economy. The estimation of consumer demand system of equations is based on the Engel curve approach described by Lyssiotou, Pashardes: food expenditure share is too low for the given income, the clothes is too high; thus the income should be corrected upwards to meet Engel curve expenditure share restrictions.

The estimates of income underreport say that both wage income and self - employment income are underreported (Table 28) (also the difference of $\theta 1$ and 1 is not significant). The average level of self - employment underreport is near 41%, while wage under - report is near 22%. The estimation method enables to use the income decomposition between wage income and self - employment income to predict the size of informal economy (the other income is supposed to be reported correctly), whereas the single equation method of Pissarides and Webber uses the arbitrarily defined type of household. The estimates of the size of informal economy may be computed for the full sample without being significantly under - estimated as in the single equation method. The value of the informal economy is near 18.8%, the figure more related to the reality. In the previous analyses even for the sample of people with positive income

from self employment the estimates account for 7.5% average level, a too moderate figure to reflect the Russian realities. The new method tells us that if we assume the differences between subsamples hold then for non couples then the size of informal income is twice as much as average for the sample, or nearly 35%. For the non couples blue collars it is near 40%.

Durable goods expenditures are very heterogeneous and tend to be reported incorrectly, that is why they are not used in the model. But the durable goods expenditures might be useful in the way that they indicate the presence of high incomes of a household, that is why the use of them will rise the black economy estimate. Whether it reflects the real economic reality is questionable.

	Sample size 1	Sample size 2
k wage_inc	1.224	1.224
k self_empl_inc	1.412	1.412
Sample size	6508	2877
Mean wage income	22720	18511
Mean self employment income	4105	9286
Mean other income	9258	9627
Mean total income	36082	37423
Mean wage income (k wage)	27813	22662
Mean income of self employed (k se)	5795	13108
Mean other income	9258	9627
Mean total income new	42866	45397
Black economy estimate	18.80%	21.31%

Table 28. The informal economy estimates based on consumer budget shares system

The quality of demand system estimation could be better if we had data available on the prices of goods purchased of a household and the time use survey to compute the value of household production. Even if the demand system is estimated without household production, the prices of aggregated goods is often needed for the own and cross price elasticities.

To conclude, we must say that methods implemented by Pissarides to the British economy and the ones implemented by Gardes, F. to the economy of Turkey have a limited application to the households of Russian Federation. Although the consumer budget share estimation seems to give the results closer to the reality due to the fact that both wage income and self - employment income tend to be under reported. The results of this system estimation are very similar to the estimates of informal economy in Turkey. Bur while estimating the size of informal economy it is necessary to mention the specific feature of Russian households statistics because of which these estimation methods have limited implementation to the database. A few of households in the sample tend to have positive income from self - employment. Thus the methods using the share of self - employment income are linked to this narrow part of database. The methods of Webber treat black economy only through the self-employment income, the simultaneous equation estimation works when this income is positive. The method is needed to estimate solely the degree of wage income underreport, to have a close look to another part of sample.

APPLICATION TABLES

	Obs	Obs>0	Total_mean	Total_sd	Total_min	Total_max
28. sausage	6517	4481	164.6	189.4	0	3000
24. pork	6517	1922	162.4	528.1	0	18000
22. beef	6517	1564	136.7	485.0	0	26100
26. bird	6517	3605	125.9	192.8	0	6000
18. fresh fruits	6517	4519	118.3	152.5	0	2080
56. tabac	6517	2548	117.1	203.8	0	3000
31. milk except dried	6517	4937	86.6	92.4	0	1200
46. fish fresh	6517	2644	86.3	200.2	0	5000
44. cookies	6517	4035	76.7	113.1	0	2038
41. chocolate	6517	3139	75.9	133.8	0	3000
1. white bread	6517	5550	63.9	61.9	0	720
36. cheese	6517	3154	62.0	93.7	0	2000
38. oil	6517	2884	45.2	83.0	0	1000
6. potatoes	6517	1749	45.1	225.0	0	5600
29. meat products	6517	1481	44.3	154.1	0	8400
32. dairy products	6517	3110	42.1	71.3	0	1000
53. vodka	6517	697	41.6	302.4	0	14576
45. eggs	6517	3718	41.1	63.6	0	2400
49. coffe	6517	1572	39.4	99.8	0	3500
40. shugar	6517	2518	38.3	124.6	0	3000
43. honey	6517	362	36.7	227.5	0	7000
34. butter	6517	2668	35.7	57.0	0	750
2. black bread	6517	4505	35.6	42.3	0	452
35. curd	6517	2373	34.7	62.6	0	600
50. soft drink	6517	1717	32.1	83.1	0	1519.1
54. vine	6517	448	31.4	390	0	28312
33. sour cream	6517	3069	30.9	46.4	0	500
55. beer	6517	1056	30.2	121.3	0	4000
3. riz	6517	2581	25.6	49.2	0	1000
5. pasta	6517	3081	25.5	40.2	0	675
48. tea	6517	2236	25.3	64.6	0	3000
25. subproducts	6517	1015	22.7	84.6	0	4000
10. tomatoes	6517	1533	19.8	46.8	0	528
4. flour	6517	1545	17.9	61.2	0	1200
20. nuts	6517	1224	17.8	59.3	0	1320
21. potted meat	6517	662	17.4	145.2	0	10000
23. lamb	6517	106	16.3	234.3	0	9999
9. cucumbers	6517	1206	14.1	38.2	0	600
12. onion	6517	1656	13.9	52.8	0	1000
8. cabbage	6517	1526	13.0	41.6	0	1000
15. melon	6517	659	12.8	53.3	0	1050
27. fat	6517	509	12.3	54.9	0	1000
37. ice cream	6517	824	11.5	40.3	0	700
17. fresh berries	6517	488	10.2	45.5	0	800
47. fish conserved	6517	733	9.3	38.9	0	1220
19. dried fruits and berries	6517	452	9.2	56.8	0	3000

51. salt	6517	1209	8.2	28.6	0	600
11. carrot	6517	1250	7.5	25.9	0	900
30. skimmed milk	6517	415	5.2	30.2	0	1200
57. bubble gum	6517	762	5.1	18.7	0	375
52. mushrooms	6517	197	5.0	58.4	0	3000
14. other vegetables	6517	334	4.3	24.2	0	400
7. canned vegetables	6517	316	3.8	22.0	0	650
13. pumpkin	6517	213	2.1	15.0	0	350
39. margarine	6517	335	1.8	13.4	0	800
16. canned berries	6517	93	1.6	16.3	0	400
42. confiture	6517	67	0.9	9.8	0	310

Table 7. Food items expenditure summary descriptive statistics grouped by mean expenditures

self_employment_income share:			
sum_D	Cum. 0	Cum. 1	Cum. 2
	0-20%	20-100%	0%
0	0.15	0.32	0.36
1	0.31	0.43	0.66
2	0.67	0.75	1.15
3	0.98	1.72	2.01
4	2.11	3.43	2.97
5	3.49	6.12	4.76
6	6.21	9.87	7.18
7	9.35	14.16	10.45
8	13.51	18.99	13.86
9	17.72	23.39	17.71
10	22.5	29.61	22.24
11	28.35	35.3	27.28
12	33.23	39.48	32.72
13	38.98	43.67	38.6
14	45.35	48.71	43.91
15	51.77	54.72	49.68
16	57.37	59.98	55.54
17	62.4	64.91	60.87
18	68	69.31	66.15
19	72.01	74.25	71.08
20	76.17	78	75.89
21	80.17	81.22	79.43
22	84.23	84.87	83.23
23	87.11	86.8	85.98
24	89.73	89.48	88.75
25	91.68	91.31	90.54
26	93.89	92.92	92.33
27	95.33	94.53	93.9
28	96.46	95.6	95.05
29	97.28	96.46	95.93

30	97.89	97.75	96.95
31	98.36	97.96	97.69
32	98.66	98.28	98.19
33	98.97	98.71	98.63
34	99.23	99.25	98.9
35	99.44	99.57	99.23
36	99.54	99.68	99.48
37	99.59	99.79	99.56
38	99.74	100	99.67
39	99.85		99.75
40	99.95		99.86
41	100		99.89
42			99.95
43			99.97
44			100
Total	1947	932	3637

Table 8. Number of food items consumed indicated by different types of households

Variable	Obs	Mean	Std. Dev.	Min	Max
clothes for adults	6508	5098	10299	0	300000
clothes for children	6508	2210	5159	0	100000
clothes_exp (90 days)	6508	7308	12619	0	300000

TV	6508	1378	6711	0	300000
mobile phone	6508	485	2475	0	70000
Furniture	6508	1362	8432	0	250000
household appliances	6508	1054	5336	0	150000
Automobile	6508	7282	74510	0	2500000
Motorcycle	6508	57	1884	0	120000
Garage	6508	112	4860	0	360000
building materials	6508	3801	29079	0	1350000
land, house, appartment	6508	4659	84689	0	2550000
Textbooks	6508	503	1303	0	20000
Bicycle	6508	95	1110	0	45000
durable_goods_exp (90 days)	6508	20789	125012	0	3493000

Variable	Obs	Mean	Std. Dev.	Min	Max
Transport	6508	665	1431	0	30000
clothes repair	6508	52	267	0	7000

TV repair	6508	34	312	0	7500
house repair	6508	1124	10574	0	400000
auto repair	6508	717	4282	0	100000
clothes wash	6508	225	762	0	35000
poste, calls	6508	77	298	0	15000
ritual service	6508	140	2229	0	60000
mobile service	6508	615	784	0	12000
internet service	6508	268	312	0	3000
lower service	6508	162	2085	0	70000
communal payment	6508	3170	3198	0	76500
service_exp (30 days)	6508	7248	13292	0	408210

Variable	Obs	Mean	Std. Dev.	Min	Max
children education	6508	411	1924	0	105000
Tourism	6508	566	5540	0	150000
Billets	6508	189	886	0	30000
health care	6508	268	2916	0	100000
health care 2	6508	168	1471	0	75000
tooth care	6508	570	4712	0	200000
Medicaments	6508	1132	1876	0	40000
Washing	6508	304	405	0	8000
private things	6508	329	473	0	6700
Cosmetics	6508	251	746	0	20000
adults courses	6508	141	1290	0	30000
Insurance	6508	247	3017	0	100100
Aliments	6508	73	861	0	30000
home tax	6508	160	1538	0	90000
health_care_exp	6508	2138	6360	0	200300
other_exp	6508	2671	8409	0	223400

Table 13. Detailed expenditure items decomposition descriptive statistics, grouped by the type of the item

REFERENCES

1. AKTUNA GUNES A., STARZEC C., GARDES F. [2011], "A new estimation of the size of informal economy using monetary and full expenditures in a complete demand system", *Paris School of Economics, University, Paris I Panthéon-Sorbonne.*
2. LYSSIOTOU P., PASHARDES P., STENGOS T. [2004], "Estimates of the Black Economy based on Consumer Demand Approaches", *The Economic Journal*, 114, p. 622-640.
3. PISSARIDES C., WEBER G. [1989], "An Expenditure-Based Estimate of Britain's Black Economy", *Journal of Public Economics*, 39, p. 17-32.
4. National Research University "Higher School of Economics" [2012], "Russian monitoring of economic conditions and health of population" (RLMS - HSE), WAVE №21.